

(Box Seq)

STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

ATTORNEYS AT LAW  
1100 NEW YORK AVENUE, N.W., SUITE 600  
WASHINGTON, D.C. 20005-3934

(202) 371-2600  
FACSIMILE (202) 371-2540, (202) 371-6566

ROBERT GREENE STERNE  
EDWARD J. KESSLER  
JORGE A. GOLDSTEIN  
SAMUEL L. FOX  
DAVID K.S. CORNWELL  
ROBERT W. ESMOND  
TRACY-GENE G. DURKIN  
MICHELE A. CIMBALA  
MICHAEL B. RAY  
ROBERT E. SOXOHL  
ERIC K. STEFFE  
MICHAEL Q. LEE

STEVEN R. LUDWIG  
JOHN M. COVERT\*  
LINDA E. ALCORN  
RAZ E. FLESHNER  
ROBERT C. MELLOWIG  
MICHAEL V. MESSINGER  
JUDITH U. KIM  
TIMOTHY J. SHEA, JR.  
DONALD R. MCPHAIL  
PATRICK E. GARRETT  
STEPHEN G. WHITESIDE  
JEFFREY T. HELVEY\*

HEIDI L. KRAUS  
JEFFREY R. KURIN  
RAYMOND MILLIEN  
PATRICK D. O'BRIEN  
LAWRENCE B. BUGAISKY  
CRYSTAL D. SAYLES\*  
EDWARD W. YEE  
ALBERT L. FERRO\*  
DONALD R. BANOWIT  
PETER A. JACKMAN  
MOLLY A. MCCALL

TERESA U. MEDLER  
JEFFREY S. WEAVER  
KRISTIN K. VIDOVICH  
KENDRICK P. PATTERSON  
DONALD J. FEATHERSTONE  
GRANT E. REED  
VINCENT L. CAPUANO  
JOHN A. HARROUN\*  
MATTHEW M. CATLETT\*  
NATHAN K. KELLEY\*  
ALBERT J. FASULO II \*

KAREN R. MARKOWICZ\*\*  
SUZANNE E. ZISKA\*\*  
BRIAN J. DEL BUONO\*\*  
ANDREA J. YAMAGUCHI\*\*  
NANCY J. LEITH\*\*  
TARJA H. NAUKKARINEN\*\*

\*BAR OTHER THAN D.C.  
\*\*REGISTERED PATENT AGENTS

10598 U.S. PTO  
09/518381  
03/03/00

March 3, 2000

WRITER'S DIRECT NUMBER:

(202) 218-7834

INTERNET ADDRESS:

BHANNES@skgf.com

10598 U.S. PTO  
03/03/00

Assistant Commissioner for Patents  
Washington, D.C. 20231

Box Patent Application

Re: U.S. Non-Provisional Utility Patent Application under 37 C.F.R. § 1.53(b)  
Appl. No. (Not Yet Assigned); Filed: Herewith  
(Cont. of 08/852,824; Filed: May 7, 1997)  
For: **Human G-Protein Coupled Receptors**  
Inventors: LI *et al.*  
Our Ref: 1488.1220002/EKS/EJH

Sir:

The following documents are forwarded herewith for appropriate action by the U.S.  
Patent and Trademark Office:

1. PTO Utility Patent Application Transmittal Form (PTO/SB/05);
2. U.S. Utility Patent Application entitled:

**Human G-Protein Coupled Receptors**

and naming as inventors:

Yi LI and  
Steven M. RUBEN

the application comprising:

Assistant Commissioner for Patents

March 3, 2000

Page 2

- a. A specification containing:
    - (i) 46 pages of description prior to the claims;
    - (ii) 3 pages of claims (1-21 claims);
    - (iii) a one (1) page abstract;
  - b. 7 sheets of drawings: (Figures 1A-1C, 2, 3A, 3B and 4);
  - c. 17 pages of a paper copy of the sequence listing;
  - d. A computer readable disk copy of the sequence listing;
  - e. A copy of the executed Declaration, as filed in U.S. Appl. No. 08/852,824;
  - f. Our check No. 26836 for \$1,974.00 to cover:  
\$ 690.00 - filing fee for patent application;  
\$1,284.00 - fee for excess claims;
3. Preliminary Amendment;
  4. Copy of a Declaration of Steven M. Ruben Under 37 C.F.R. § 1.132 with attachments A-C filed in prosecution of U.S. Appl. No. 08/852,824;
  5. PTO Fee Transmittal Form PTO/SB/17 (*in duplicate*);
  6. Authorization to Treat a Reply As Incorporating An Extension of Time Under 37 C.F.R. § 1.136(a)(3) (*in duplicate*); and
  7. Two (2) return postcards.

In accordance with 37 C.F.R. § 1.821(f), the paper copy of the sequence listing and the computer readable copy of the sequence listing submitted herewith are the same.

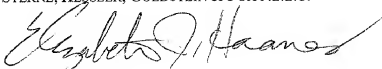
Assistant Commissioner for Patents  
March 3, 2000  
Page 3

It is respectfully requested that, of the two attached postcards, one be stamped with the filing date of these documents and returned to our courier, and the other, prepaid postcard, be stamped with the filing date and unofficial application number and returned as soon as possible.

The U.S. Patent and Trademark Office is hereby authorized to charge any fee deficiency, or credit any overpayment, to our Deposit Account No. 19-0036. A duplicate copy of this letter is enclosed.

Respectfully submitted,

STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.



Elizabeth J. Haanes  
Agent for Applicants  
Registration No. 42,613

EKS/EJH/cmg  
Enclosures

P:\USERS\BHAANES\Work Products\1488\122\1220002\CONT-tml



Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

## UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR § 1.53(b))

Attorney Docket No.	1488.1220002/EKS/EJH
First Inventor or Applicant Identifier	LI et al
Title	Human G-Protein Coupled Receptors
Express Mail Label No	

## APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents

ADDRESS TO

Assistant Commissioner for Patents  
Box Patent Application  
Washington, DC 20231

1. ☒ \* Fee Transmittal Form (e.g., PTO/SB/17)  
(Submit an original, and a duplicate for fee processing)

6. ☐ Microfiche Computer Program (Appendix)

2. ☒ Specification [Total Pages 50 ]

(preferred arrangement set forth below)

- Descriptive title of the Invention
- Cross References to Related Applications
- Statement Regarding Fed-sponsored R & D
- Reference to Microfiche Appendix
- Background of the Invention
- Brief Summary of the Invention
- Brief Description of the Drawings (if filed)
- Detailed Description
- Claim(s)
- Abstract of the Disclosure

7. Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)

- a. ☒ Computer Readable Copy  
b. ☒ Paper Copy (identical to computer copy)  
c. ☒ Statement verifying identity of above copies

3. ☒ Drawing(s) (35 U.S.C. 113) [Total Sheets 7 ]

4. ☒ Oath or Declaration [Total Pages 1 ]

- a. ☐ Newly executed (original or copy)

- b. ☒ Copy from a prior application (37 CFR 1.63(d)) (for continuation/divisional with Box 17 completed) [Note Box 5 below]

- i. ☐ DELETION OF INVENTOR(S)  
Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR §§ 1.63(d)(2) and 1.33(b).

5. ☒ Incorporation By Reference (useable if Box 4b is checked)  
The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein

## ACCOMPANYING APPLICATION PARTS

8. ☐ Assignment Papers (cover sheet & document(s))  
9. ☐ 37 CFR 3.73(b) Statement ☐ Power of Attorney (when there is an assignee)  
10. ☐ English Translation Document (if applicable)  
11. ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations  
12. ☒ Preliminary Amendment  
13. ☒ Return Receipt Postcard (MPEP 503) (Should be specifically itemized)  
14. ☐ \*Small Entity Statement(s) ☐ Statement filed in prior application, Status still proper and desired (PTO/SB/09-12)  
15. ☐ Certified Copy of Priority Document(s) (if foreign priority is claimed)  
16. ☒ Other 37 C.F.R. § 1.136(a)(3) Authorization

☒ Other Copy of Declaration of Steven M. Ruben Under 37 C.F.R. § 1.132 w/attachments A-C

\*NOTE FOR ITEMS 1 & 14: IN ORDER TO BE ENTITLED TO PAY SMALL ENTITY FEES, A SMALL ENTITY STATEMENT IS REQUIRED (37 C.F.R. § 1.17), EXCEPT IF ONE FILED IN A PRIORITY APPLICATION IS RELIED UPON (37 C.F.R. § 1.18)

17. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment

- ☒ Continuation ☐ Divisional ☐ Continuation-in-Part (CIP) of prior application No: 08/852,824

Prior application information: Examiner: N. Basi Group/Art Unit: 1646

## 18. CORRESPONDENCE ADDRESS

☐ Customer Number  
or Bar Code Label

(Insert Customer No. or Attach bar code label here)

or ☒ Correspondence  
address below

NAME	STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.				
	Attorneys at Law				
ADDRESS	Suite 600, 1100 New York Avenue, N.W.				
CITY	Washington	STATE	DC	ZIP CODE	20005-3934
COUNTRY	USA	TELEPHONE	(202) 371-2600	FAX	(202) 371-2540

NAME (Print/Type)	Elizabeth J. Haagen	Registration No. (Attorney/Agent)	42,613
SIGNATURE	<i>Elizabeth J. Haagen</i>	Date	March 3, 2000

Burden Hour Statement: this form is estimated to take 02 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Li *et al.*

Appl. No.: To Be Assigned  
(Cont. of Appl. No. 08/852,824)

Filed: Herewith

For: **Human G-Protein Coupled  
Receptors**

Art Unit: To Be Assigned

Examiner: To Be Assigned

Atty. Docket: 1488.1220002/EKS/EJH

**Preliminary Amendment**

Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

In advance of prosecution, please amend the application as follows.

***In the Claims***

Please cancel claims 1-21 (22 total claims, two claims were inadvertently listed as claim 4) without prejudice to or disclaimer of the subject matter thereof.

Please add the following claims 23-96:

~~-- 23.~~ An isolated polynucleotide comprising a nucleic acid at least 90% identical to 30 contiguous nucleotides of SEQ ID NO:3.

24. The polynucleotide of claim 23, wherein said nucleic acid is at least 95% identical to 30 contiguous nucleotides of SEQ ID NO:3.

25. The polynucleotide of claim 24, comprising 30 contiguous nucleotides of SEQ ID NO:3.
26. The polynucleotide of claim 23, further comprising a heterologous polynucleotide.
27. A vector comprising the polynucleotide of claim 23.
28. A host cell comprising the polynucleotide of claim 23.
29. The host cell of claim 28, wherein said polynucleotide is operably associated with a heterologous regulatory sequence.
30. A method of using the host cell of claim 29 to screen for ligand binding, comprising culturing said host cell under conditions such that a polypeptide encoded by said polynucleotide is expressed, contacting said polypeptide with said ligand, and detecting binding of said ligand to said polypeptide.
31. A method of producing a polypeptide comprising culturing the host cell of claim 29 under conditions such that said polypeptide is expressed, and recovering said polypeptide.
32. A polypeptide produced by the method of claim 31.

33. An isolated polynucleotide comprising a nucleic acid at least 90% identical to 50 contiguous nucleotides of SEQ ID NO:3.

34. The polynucleotide of claim 33, wherein said nucleic acid is at least 95% identical to 50 contiguous nucleotides of SEQ ID NO:3.

35. The polynucleotide of claim 34, comprising 50 contiguous nucleotides of SEQ ID NO:3.

36. The polynucleotide of claim 33, further comprising a heterologous polynucleotide.

37. A vector comprising the polynucleotide of claim 33.

38. A host cell comprising the polynucleotide of claim 33.

39. The host cell of claim 38, wherein said polynucleotide is operably associated with a heterologous regulatory sequence.

40. A method of using the host cell of claim 39 to screen for ligand binding, comprising culturing said host cell under conditions such that a polypeptide encoded by said polynucleotide is expressed, contacting said polypeptide with said ligand, and detecting binding of said ligand to said polypeptide.

41. A method of producing a polypeptide comprising culturing the host cell of claim 39 under conditions such that said polypeptide is expressed, and recovering said polypeptide.

42. A polypeptide produced by the method of claim 41.

43. An isolated polynucleotide comprising a nucleic acid at least 90% identical to 150 contiguous nucleotides of SEQ ID NO:3.

44. The polynucleotide of claim 43, wherein said nucleic acid is at least 95% identical to 150 contiguous nucleotides of SEQ ID NO:3.

45. The polynucleotide of claim 44, comprising 150 contiguous nucleotides of SEQ ID NO:3.

46. The polynucleotide of claim 43, further comprising a heterologous polynucleotide.

47. A vector comprising the polynucleotide of claim 43.

48. A host cell comprising the polynucleotide of claim 43.

49. The host cell of claim 48, wherein said polynucleotide is operably associated with a heterologous regulatory sequence.



50. A method of using the host cell of claim 49 to screen for ligand binding, comprising culturing said host cell under conditions such that a polypeptide encoded by said polynucleotide is expressed, contacting said polypeptide with said ligand, and detecting binding of said ligand to said polypeptide.

51. A method of producing a polypeptide comprising culturing the host cell of claim 49 under conditions such that said polypeptide is expressed, and recovering said polypeptide.

52. A polypeptide produced by the method of claim 51.

53. An isolated polynucleotide comprising a nucleic acid at least 90% identical to a reference nucleic acid encoding 30 contiguous amino acids of SEQ ID NO:4.

54. The polynucleotide of claim 53, wherein said nucleic acid is at least 95% identical to said reference nucleic acid.

55. The polynucleotide of claim 54, wherein said nucleic acid encodes 30 contiguous amino acids of SEQ ID NO:4.

56. The polynucleotide of claim 53, wherein said nucleic acid encodes a polypeptide which binds an antibody having specificity for the polypeptide of SEQ ID NO:4.

57. The polynucleotide of claim 53, wherein said nucleic acid encodes a polypeptide which as G protein-coupled receptor activity.

58. The polynucleotide of claim 53, further comprising a heterologous polynucleotide.

59. A vector comprising the polynucleotide of claim 53.

60. A host cell comprising the polynucleotide of claim 53.

61. The host cell of claim 60, wherein said polynucleotide is operably associated with a heterologous regulatory sequence.

62. A method of using the host cell of claim 61 to screen for ligand binding, comprising culturing said host cell under conditions such that a polypeptide encoded by said polynucleotide is expressed, contacting said polypeptide with said ligand, and detecting binding of said ligand to said polypeptide.

63. A method of producing a polypeptide comprising culturing the host cell of claim 61 under conditions such that said polypeptide is expressed, and recovering said polypeptide.

64. A polypeptide produced by the method of claim 63.

65. An isolated polynucleotide comprising a nucleic acid at least 90% identical to a reference nucleic acid encoding 50 contiguous amino acids of SEQ ID NO:4.

66. The polynucleotide of claim 65, wherein said nucleic acid is at least 95% identical to said reference nucleic acid.

67. The polynucleotide of claim 66, wherein said nucleic acid encodes 50 contiguous amino acids of SEQ ID NO:4.

68. The polynucleotide of claim 65, wherein said nucleic acid encodes a polypeptide which binds an antibody having specificity for the polypeptide of SEQ ID NO:4.

69. The polynucleotide of claim 65, wherein said nucleic acid encodes a polypeptide which has G protein-coupled receptor activity.

70. The polynucleotide of claim 65, further comprising a heterologous polynucleotide.

71. A vector comprising the polynucleotide of claim 65.

72. A host cell comprising the polynucleotide of claim 65.

73. The host cell of claim 72, wherein said polynucleotide is operably associated with a heterologous regulatory sequence.

74. A method of using the host cell of claim 73 to screen for ligand binding, comprising culturing said host cell under conditions such that a polypeptide encoded by said polynucleotide is expressed, contacting said polypeptide with said ligand, and detecting binding of said ligand to said polypeptide.

75. A method of producing a polypeptide comprising culturing the host cell of claim 73 under conditions such that said polypeptide is expressed, and recovering said polypeptide.

76. A polypeptide produced by the method of claim 75.

~~77.~~ An isolated polynucleotide comprising a nucleic acid encoding at least one transmembrane domain of SEQ ID NO:4.

78. The polynucleotide of claim 77, further comprising a heterologous polynucleotide.

79. A vector comprising the polynucleotide of claim 77.

80. A host cell comprising the polynucleotide of claim 77.

81. The host cell of claim 80, wherein said polynucleotide is operably associated with a heterologous regulatory sequence.

82. A method of using the host cell of claim 81 to screen for ligand binding, comprising culturing said host cell under conditions such that a polypeptide encoded by said polynucleotide is expressed, contacting said polypeptide with said ligand, and detecting binding of said ligand to said polypeptide.

83. A method of producing a polypeptide comprising culturing the host cell of claim 81 under conditions such that said polypeptide is expressed, and recovering said polypeptide.

84. A polypeptide produced by the method of claim 83.

~~85.~~ An isolated polypeptide comprising amino acids, wherein the sequence of said amino acids is at least 90% identical to 30 contiguous amino acids of SEQ ID NO:4.

86. The polypeptide of claim 85, wherein the sequence of said amino acids is at least 95% identical to 30 contiguous amino acids of SEQ ID NO:4.

87. The polypeptide of claim 86, comprising 30 contiguous amino acids of SEQ ID NO:4.

88. The polypeptide of claim 85, comprising amino acids, wherein the sequence of said amino acids is at least 90% identical to 50 contiguous amino acids of SEQ ID NO:4.

89. The polypeptide of claim 88, wherein the sequence of said amino acids is at least 95% identical to 50 contiguous amino acids of SEQ ID NO:4.

90. The polypeptide of claim 89, comprising 50 contiguous amino acids of SEQ ID NO:4.

91. The polypeptide of claim 85, wherein said polypeptide has G-protein coupled receptor activity.

92. The polypeptide of claim 85, wherein said polypeptide binds an antibody having specificity for the polypeptide of SEQ ID NO:4.

93. The polypeptide of claim 85, further comprising a heterologous polypeptide.

94. A composition comprising the polypeptide of claim 85 and a carrier.

95. An antibody which binds the polypeptide of claim 85.

96. An antagonist of the polypeptide of claim 85.--

***Remarks***

New claims 23-96 have been added to more precisely define the invention. Support for added claims 23-96 may be found throughout the specification, in Figures 2 and 4, and in the attached copy of a Declaration under 37 C.F.R. § 1.132 by Steven M. Ruben (the Ruben Declaration) filed in conjunction with an Amendment and Reply filed with the parent case, Application No. 08/852,824 (the '824 Application).

Applicants hereby state that no new matter has been added by way of the foregoing amendments.

***The Sequence Listing***

The Sequence Listing submitted herewith corrects typographical errors present in the Sequence Listing originally filed with the '824 Application. The undersigned hereby states that the changes made in the Sequence Listing do not include new matter.

In accordance with 37 C.F.R. § 1.825(b), the undersigned hereby states that the paper copy of the Sequence Listing and the computer readable copy of the Sequence Listing submitted herewith are the same.

Typographical errors were incorporated in the Sequence Listing originally filed with the '824 Application due to attorney error. The errors appeared in SEQ ID NO:1 through SEQ ID NO:4 of the Sequence Listing, and since copies of the Sequence Listing pages for SEQ ID NO:1 and SEQ ID NO:3 were used as informal drawings in Figures 1 and 3 as originally filed, the errors also appeared in those drawings. Applicants stress that these errors were incorporated during the drafting of the application, and that Applicants were in possession of the correct sequence at the

time the application was filed. Support may be found, *inter alia*, in Figures 2 and 4, and in the attached copy of the Ruben Declaration. In particular, SEQ ID NO:1 as originally filed in the '824 Application contained typographical errors at the following nucleotide positions: position 242, which should be A rather than T; position 266, which should be C rather than A; position 1870, where a T should be deleted, and position 2206, where an N should be deleted. SEQ ID NO:2 as originally filed in the '824 Application contained typographical errors at the following amino acid positions: position 6, which should be Asn rather than Ile; and position 14, which should be Thr rather than Asn. The latter typographical errors in the amino acid sequence were also reflected in the amino acid sequence shown under the nucleotide sequence in SEQ ID NO:1 as originally filed in the '824 Application. SEQ ID NO:3 as originally filed in the '824 Application contained typographical errors at the following nucleotide positions: position 828, which should be T rather than C; and position 831, which should be T rather than A. Note that this latter typographical error introduced a stop codon into the open reading frame, causing the amino acid sequence, as translated from the sequence with the typographical error, to stop at position 260. SEQ ID NO:4, therefore, was truncated at position 260, which should be Phe rather than Ser, and the translation should continue to amino acid 384. SEQ ID NO:4 as originally filed in the '824 Application contained typographical errors at the following amino acid positions: position 191, which should be Asp rather than Asn, position 202, which should be Lys rather than Arg, and position 204, which should be Tyr rather than Thr. In addition, the translation should start with the Met at position 1, rather than the Ala at position -16. The latter typographical errors in the amino acid sequence were also reflected in the amino acid sequence shown under the nucleotide sequence in SEQ ID NO:3 as originally filed in the '824 Application.



In support of the fact that Applicants indeed had the correct sequence at the time the '824 Application was filed, the sequence of SEQ ID NO:2, as amended, is shown in the top line of the alignment shown in Fig. 2, and the sequence of SEQ ID NO:4, as amended, is shown in the top line of the alignment shown in Fig. 4, except for five residues at the 3' end of the polypeptide. These latter five residues are not in the alignment simply because they did not align with the second sequence, *i.e.*, SEQ ID NO:18.

Claims 23-84 are directed to a polynucleotide. In other words, these claims are directed to a chemical compound and not to a nucleotide sequence. The nucleotide sequence is but an inherent property of the claimed polynucleotide.

There is a line of chemical case law where applicants have been permitted to amend the specification to correct the formula of a chemical compound after an application's filing date provided that it had been demonstrated that one of skill in the art would have appreciated that the applicant was in possession of the compound itself at the time of filing. The rationale is that the formula is an inherent property of the compound and thus amending the specification to correct the formula is not new matter. See *In re Nathan*, 140 U.S.P.Q. 601, 604 (C.C.P.A. 1964). *Accord Kennecott Corp. v. Kyocera Int'l, Inc.*, 5 U.S.P.Q.2d 1194, 1198 (Fed. Cir. 1987), *cert. denied*, 486 U.S. 1008 (1988) ("The disclosure in a subsequent patent application of an inherent property of a product does not deprive that product of the benefit of the earlier filing date.").

In the field of biotechnology, applicants often rely on a deposited clone, where the deposit was made prior to filing, to establish possession of claimed polynucleotides or proteins. The focus for determining whether applicants were in possession of these polynucleotides or proteins has been determined, at least in part, by considering whether the applicant has (1) established that one skilled in the art in possession of the deposited clone would have been aware of both the DNA

sequence and the encoded amino acid sequence, or would be able to determine these sequences without undue experimentation, (2) established that the DNA and amino acids sequences are described in a manner such that one skilled in the art could distinguish them from other sequences, and (3) resequenced a clone which is identical to that of the deposit and established a "chain of custody" for this clone. See e.g., *Ex parte Maizel*, 27 U.S.P.Q.2d 1662, 1669-1670 (B.P.A.I. 1992).

At paragraphs 8 and 14 of the attached copy of the Ruben Declaration, Dr. Ruben states that he is of the opinion that the correct EBI-2 and EDG-1-like nucleotide and amino acid sequences would have been apparent, as of the May 7, 1997 filing date of the '824 Application, to one skilled in the art in possession of ATCC Deposit Nos. 209003 and 209004 and the sequence data originally filed with the present application. Dr. Ruben bases this position on the facts that the correct EBI-2 and EDG-1-like coding sequences can be readily determined from the deposited clones, and methods for sequencing these clones were routine in the art in May of 1997 (Ruben Declaration, paras. 8 and 14). Thus, the nucleotide sequences of these clones are chemical structures which are inherent properties of the clones. Further, the EBI-2 and EDG-1-like sequences are described in the present application in sufficient detail so that one skilled in the art could distinguish the claimed invention from other compounds.

The typographical errors present in the Sequence Listing as filed were incorporated into the '824 Application during drafting by the attorneys, and were not the errors of the Applicants. Indeed, at the time of filing of the '824 Application, Applicants possessed the nucleotide sequences SEQ ID NO:1 and SEQ ID NO:3 (as amended) of the two cDNA clones (HHPGS02 and HNFDL69) that encode the EBI-2 and EDG-1-like amino acid sequences of SEQ ID NO:2 and SEQ ID NO:4 (as amended) of the present application (Ruben Declaration, paras. 6-7 and

12-13). Applicants have also submitted evidence to establish that these clones are identical to those deposited at the ATCC and given Accession Nos. 209003 and 209004 (Ruben Declaration, paras. 4 and 10). Applicants have thus established a "chain of custody" which demonstrates that the nucleotide sequence data included in SEQ ID NO: 1 and SEQ ID NO:3 (as amended) of the present application was obtained from the clones deposited with the ATCC and given Accession Nos. 209003 and 209004.

In light of this information, the inclusion of the correct sequences present in the clones of ATCC Deposit Nos. 209003 and 209004, and in possession of the Applicants at the time of filing of the '824 Application, does not represent new matter.

***Summary***

It is respectfully believed that this application is now in condition for examination. Early notice to this effect is respectfully requested

The U.S. Patent and Trademark Office is hereby authorized to charge any fee deficiency, or credit any overpayment, to our Deposit Account No. 19-0036.

Respectfully submitted,

STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.



Elizabeth J. Haanes  
Agent for Applicant  
Registration No. 42,613

Date: March 3, 2000

1100 New York Avenue, N.W.  
Suite 600  
Washington, D.C. 20005  
(202) 371-2600  
P94-34.wpd

## ***HUMAN G-PROTEIN COUPLED RECEPTORS***

This Application is a continuation of U.S. Application no. 08/852,824, filed May 7, 1997, the disclosure of which is incorporated herein by reference in its entirety.

### ***Background of the Invention***

#### ***Field of the Invention***

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides. More particularly, the polypeptides of the present invention are a human EBV-induced G-protein coupled receptor (EBI-2) and a human EDG-1-like G-protein coupled receptor, sometimes hereinafter referred to singularly as "GPR" or "GPCR" and collectively as "GPRs." The invention also relates to inhibiting the action of such polypeptides.

#### ***Related Art***

At least nine genes have been identified that are apparently activated in response to an Epstein-Barr Virus (EBV) infection. One of two novel genes also identified in such studies of EBV infections was a novel GPCR-like cDNA molecule designated EBV-induced G-protein coupled receptor (EBI)-1.

Additionally, previously identified was an endothelium-differentiation gene (EDG) that was obtained from PMA-simulated human endothelial cells. Rat and sheep homologs of EDG-1 have been identified, which are also G-protein coupled receptors.

It is well established that many medically significant biological processes are mediated by proteins participating in signal transduction pathways that involve G-proteins and/or second messengers, *e.g.*, cAMP (Lefkowitz, *Nature* 351:353-354 (1991)). Herein these proteins are referred to as proteins participating in pathways-with G-proteins or PPG proteins. Some examples of these proteins include the GPC receptors, such as those for adrenergic agents and dopamine

For example, in one form of signal transduction, the effect of hormone binding is activation of an enzyme, adenylate cyclase, inside the cell. Enzyme activation by hormones is dependent on the presence of the nucleotide GTP, and GTP also influences hormone binding. G-protein connects the hormone receptors to adenylate cyclase. G-protein was shown to exchange GTP for bound GDP when activated by hormone receptors. The GTP-carrying form then binds to an activated adenylate cyclase. Hydrolysis of GTP to GDP, catalyzed by the G-protein itself, returns the G-protein to its basal, inactive form. Thus, the G-protein serves a dual role, as an intermediate that relays the signal from receptor to effector, and as a clock that controls the duration of the signal.

G-protein coupled receptors have been characterized as including these seven conserved hydrophobic stretches of about 20 to 30 amino acids, connecting

at least eight divergent hydrophilic loops. The G-protein family of coupled receptors includes dopamine receptors which bind to neuroleptic drugs used for treating psychotic and neurological disorders. Other examples of members of this family include calcitonin, adrenergic, endothelin, cAMP, adenosine, muscarinic, acetylcholine, serotonin, histamine, thrombin, kinin, follicle stimulating hormone, opsins and rhodopsins, odorant, cytomegalovirus receptors, etc.

Most GPRs have single conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that are believed to stabilize functional protein structure. The 7 transmembrane regions are designated as TM1, TM2, TM3, TM4, TM5, TM6, and TM7. TM3 is also implicated in signal transduction.

Phosphorylation and lipidation (palmitoylation or farnesylation) of cysteine residues can influence signal transduction of some GPRs. Most GPRs contain potential phosphorylation sites within the third cytoplasmic loop and/or the carboxy terminus. For several GPRs, such as the  $\beta$ -adrenoreceptor, phosphorylation by protein kinase A and/or specific receptor kinases mediates receptor desensitization.

The ligand binding sites of GPRs are believed to comprise a hydrophilic socket formed by several GPR transmembrane domains, which socket is surrounded by hydrophobic residues of the GPRs. The hydrophilic side of each GPR transmembrane helix is postulated to face inward and form the polar ligand binding site. TM3 has been implicated in several GPRs as having a ligand binding site, such as including the TM3 aspartate residue. Additionally, TM5 serines, a TM6 asparagine and TM6 or TM7 phenylalanines or tyrosines are also implicated in ligand binding.

GPRs can be intracellularly coupled by heterotrimeric G-proteins to various intracellular enzymes, ion channels and transporters (see, Johnson *et al.*, *Endoc. Rev.* 10:317-331 (1989)). Different G-protein  $\alpha$ -subunits preferentially stimulate particular effectors to modulate various biological functions in a cell. Phosphorylation of cytoplasmic residues of GPRs has been identified as an important mechanism for the regulation of G-protein coupling of some GPRs.

G-protein coupled receptors are found in numerous sites within a mammalian host. For example, dopamine is a critical neurotransmitter in the central nervous system and is a G-protein coupled receptor ligand.

### *Summary of the Invention*

In accordance with one aspect of the present invention, there are provided novel polypeptides, as well as antisense analogs thereof and biologically active and diagnostically or therapeutically useful fragments and derivatives thereof. The polypeptides of the present invention are of human origin.

In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules, including mRNAs, DNAs, cDNAs, genomic DNA as well as antisense analogs thereof and biologically active and diagnostically or therapeutically useful fragments thereof.

In accordance with a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques which comprises culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence encoding a polypeptide of the present invention, under conditions promoting expression of said protein and subsequent recovery of said protein.

In accordance with yet a further aspect of the present invention, there are provided antibodies against such polypeptides.

In accordance with another embodiment, there is provided a process for using one or more of the receptors according to the invention to screen for receptor antagonists and/or agonists and/or receptor ligands.

In accordance with still another embodiment of the present invention there is provided a process of using such agonists to activate the polypeptide of the present invention for the treatment of conditions related to the underexpression of the polypeptide of the present invention.

In accordance with another aspect of the present invention there is provided a process of using such antagonists for inhibiting the polypeptide of the



present invention for treating conditions associated with overexpression of the polypeptide of the present invention.

In accordance with yet another aspect of the present invention there is provided non-naturally occurring synthetic, isolated and/or recombinant polypeptides which are fragments, consensus fragments and/or sequences having conservative amino acid substitutions, of at least one transmembrane domain, such that the polypeptides of the present invention may bind ligands, or which may also modulate, quantitatively or qualitatively, ligand binding to the polypeptide of the present invention.

In accordance with still another aspect of the present invention there are provided synthetic or recombinant polypeptides, conservative substitution derivatives thereof, antibodies, anti-idiotypic antibodies, compositions and methods that can be useful as potential modulators of G-protein coupled receptor function, by binding to ligands or modulating ligand binding, due to their expected biological properties, which may be used in diagnostic, therapeutic and/or research applications.

In accordance with another object of the present invention, there is provided synthetic, isolated or recombinant polypeptides which are designed to inhibit or mimic various GPRs or fragments thereof, as receptor types and subtypes.

In accordance with yet another object of the present invention, there is provided a diagnostic assay for detecting a disease or susceptibility to a disease related to a mutation in a nucleic acid sequence encoding a polypeptide of the present invention.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

### ***Brief Description of the Figures***

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figures 1A, 1B, and 1C show the cDNA sequence (SEQ ID NO:1) and the corresponding deduced amino acid sequence (SEQ ID NO:2) of the EBV-induced G-protein coupled receptor of the present invention. The polynucleotide sequence contains a 2249 nucleotide sequence which encodes a 342 amino acid ORF. In Figures 1A-1C, the standard one-letter abbreviation for amino acids is used to illustrate the deduced amino acid sequence. Sequencing was performed using a 373 Automated DNA sequencer (Applied Biosystems, Inc.). Sequencing accuracy is predicted to be greater than 97% accurate.

Figure 2 is an amino acid sequence comparison between the EBV-induced (EBI-2) G-Protein Coupled Receptor (upper line, see SEQ ID NO:2) and the human EBI-1 G-Protein Coupled Receptor (lower line, SEQ ID NO: 17). The standard one letter abbreviations are used to represent the amino acid residues of the amino acid sequences illustrated. The EBI-2 polypeptide according to the invention shows approximately 25% identity and 49% similarity to the amino acid sequence of the EBI-1 gene over an approximately 350 amino acid stretch.

Figures 3A and 3B show the cDNA sequence (SEQ ID NO:3) and the corresponding deduced- amino acid sequence (SEQ ID NO:4) of the EDG-1-like G-protein coupled receptor of the present invention. The polynucleotide sequence contains a 1637 nucleotide sequence which encodes a 384 amino acid ORF. In Figures 3A and 3B, the standard one-letter abbreviation for amino acids is used to illustrate the deduced amino acid sequence. Sequencing was performed using a 373 Automated DNA sequencer (Applied Biosystems, Inc.). Sequencing accuracy is predicted to be greater than 97% accurate.

Figure 4 is an amino acid sequence comparison between the EDG-1-like G-Protein Coupled Receptor (upper line, see SEQ ID NO:4) and the human EDG-1 orphan G-Protein Coupled Receptor (lower line, SEQ ID NO:18). The standard one-letter abbreviations are used to represent the amino acid residues of the amino acid sequences illustrated. The EDG-1-like polypeptide according to the invention shows approximately 54% identity and 73% similarity to the amino acid sequence of the human EDG-1 orphan G-protein Coupled Receptor gene over two regions totaling approximately 120 amino acids.

### ***Detailed Description of the Invention***

In accordance with an aspect of the present invention, there is provided an isolated nucleic acid (polynucleotide) which encode for the mature polypeptide having the deduced amino acid sequence of Figures 1A-1C (SEQ ID NO: 2) or for the

mature polypeptide encoded by the cDNA of the clone deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, as ATCC Deposit No. 209003 on 4/28/97.

A polynucleotide encoding an EBI-2 polypeptide of the present invention may be found in a cDNA library from umbilical vein endothelial cells, neutrophil leukocyte cells, and corpus colosum cells. The polynucleotide of this invention was discovered in a cDNA library derived from umbilical vein endothelial cells. As described above, it is structurally related to the G protein-coupled receptor family. It contains an open reading frame encoding a protein of 343 amino acid residues.

In accordance with an aspect of the present invention, there is provided an isolated nucleic acid (polynucleotide) which encodes for the mature polypeptide having the deduced amino acid sequence of Figures 3A and 3B (SEQ ID NO:4) or for the mature polypeptide encoded by the cDNA of the clone deposited as ATCC Deposit No. 209004 on 4/28/97.

A polynucleotide encoding an EDG-1-like G-protein coupled receptor polypeptide of the present invention may be found in an activated neutrophil cDNA library, cyclohexamine-treated Raji cells, the RSR;11 bone marrow cell line, activated T-cells, tonsils, and CD34-positive cord blood cells. Northern blot analyses indicate that the EDG-1-like receptor gene is expressed primarily in leukocytes, but expression may also be observed in placenta, spleen, thymus, lung and pancreas tissue. The polynucleotide of this invention was discovered in a cDNA library derived from activated neutrophils. As described above, it is structurally related to the G protein-coupled receptor family. It contains an open reading frame encoding a protein of 384 amino acid residues.

As noted above, a great deal of the importance attributed to GPCR molecules such as those of the presently claimed invention lies in the diversity of biological functions in which they participate. For example, it is thought that, upon release from the alpha subunit, the beta/gamma subunit may also play a functional role in the regulation of signal transduction by activating the arachidonic acid signal transduction pathway via the activation of phospholipase A<sub>2</sub>. In addition, GPCR molecules and their associated G-proteins have been implicated in the coupling of visual pigments to cGMP phosphodiesterase, phosphatidyl inositol (PI) turnover, adenylyl cyclase signal channels and other integral membrane enzymes to transporter proteins. As a result, it is apparent that novel GPCR molecules may prove useful in a wide variety of pharmaceutical applications including research and development. For example, target based screens for small molecules and other such pharmacologically valuable factors may be based on activating a given GPCR. It has also been observed that short peptides may function by mimicking the GPCR (termed receptomimetics). Furthermore, monoclonal antibodies raised against such factors may prove useful as therapeutics in a number of capacities. Potential therapeutic and/or diagnostic applications for such a factor may include such diverse clinical presentations as heart disease, mental illness, cancer, atherosclerosis, restenosis, Alzheimer's Disease, Parkinson's Disease, and a number of others.

Accordingly, the polynucleotides of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the mature EBI-2 polypeptide may be identical to the coding sequence shown in Figures 1A-1C (SEQ ID NO:1) or that of the deposited clone or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature polypeptide as the DNA of Figures 1A-1C (SEQ ID NO:1) or the deposited cDNA. Similarly, the coding sequence which encodes the mature EDG-1-like G-protein coupled receptor polypeptide may be identical to the coding sequence shown in Figures 3A and 3B (SEQ ID NO: 3) or that of the deposited

The polynucleotides which encode either (a) the mature EBI-2 polypeptide of Figures 1A-1C (SEQ ID NO:2) or the mature EBI-2 polypeptide encoded by the deposited cDNA, or (b) the mature EDG-1-like G-protein coupled receptor polypeptide of Figures 3A and 3B (SEQ ID NO:4) or the mature EDG-1-like G-protein coupled receptor polypeptide encoded by the deposited cDNA may include: only the coding sequence for the mature polypeptide; the coding sequence for the mature polypeptide and additional coding sequence such as a leader or secretory sequence or a proprotein sequence; the coding sequence for the mature polypeptide (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature polypeptide.

Thus, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of (a) the polypeptide having the deduced amino acid sequence of Figures 1A-1C (SEQ ID NO:2) or the polypeptide encoded by the cDNA of the deposited clone, or (2) the polypeptide having the deduced amino acid sequence of Figures 3A and 3B (SEQ ID NO:4) or the polypeptide encoded by the cDNA of the deposited clone. The variant of either of these two polynucleotides may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature polypeptide as shown in Figures 1A-1C (SEQ ID NO:2) or the same mature polypeptide encoded by the cDNA of the deposited clone as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the polypeptide of Figures 1A-1C (SEQ ID NO:2) or the polypeptide

encoded by the cDNA of the deposited clone. Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

Likewise, the present invention includes polynucleotides encoding the same mature polypeptide as shown in Figures 3A and 3B (SEQ ID NO:4) or the same mature polypeptide encoded by the cDNA of the deposited clone as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the polypeptide of Figures 3A and 3B (SEQ ID NO:4) or the polypeptide encoded by the cDNA of the deposited clone. Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

As hereinabove indicated, the polynucleotide may have a coding sequence which is a naturally occurring allelic variant of the coding sequence shown in Figures 1A-1C (SEQ ID NO:1) or of the coding sequence of the deposited clone. Also, as hereinabove indicated, the polynucleotide may have a coding sequence which is a naturally occurring allelic variant of the coding sequence shown in Figures 3A and 3B (SEQ ID NO:3) or of the coding sequence of the deposited clone. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

The present invention also includes polynucleotides, wherein the coding sequence for the mature polypeptide may be fused in the same reading frame to a polynucleotide sequence which aids in expression and secretion of a polypeptide from a host cell, for example, a leader sequence which functions as a secretory sequence for controlling transport of a polypeptide from the cell. The polypeptide having a leader sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature form of the polypeptide. The polynucleotides may also code for a proprotein which is the mature protein plus additional 51 amino acid residues. A mature protein having a prosequence is a proprotein and is an inactive form of the protein. Once the prosequence is cleaved an active mature protein remains.

Thus, for example, the polynucleotide of the present invention may encode a mature protein, or a protein having a prosequence or for a protein having both a prosequence and a presequence (leader sequence).

The polynucleotides of the present invention may also have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptide of the present invention. The marker sequence may be a hexahistidine tag supplied by a pQE-9 vector to provide for purification of the mature polypeptide fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, *e.g.* COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., *et al.*, *Cell* 37:767 (1984)).

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

Fragments of the full length gene of the present invention may be used as hybridization probes for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type preferably have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. In fact, probes of this type having at least up to 150 bases or greater may be preferably utilized. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promoter regions, exons and introns. An example of a screen comprises isolating the coding region of the gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary or identical to that of the gene or portion of the gene sequences of the present invention are used to screen a library of genomic DNA to determine which members of the library the probe hybridizes to.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe.

Useful reagents include but are not limited to radioactivity, fluorescent dyes, or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate complementary copies of DNA from other sources or to screen such sources for related sequences.

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. (As indicated above, 70% identity would include within such definition a 70 bps fragment taken from a 100 bp polynucleotide, for example.) The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode enzymes which either retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA of Figures 1 A-C, 3A and 3B (SEQ ID NOS:2 and 4, respectively). In referring to identity in the case of hybridization, as known in the art, such identity refers to complementarity of polynucleotide segments.

Alternatively, the polynucleotide may have at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to any part of a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed as probes for the polynucleotides of SEQ ID NOS:1 and 3, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity, preferably at least 90% identity and more preferably at least a 95% identity to a polynucleotide which encodes either the polypeptide of SEQ ID NO:2, or the polypeptide of SEQ ID NO:4, as well as fragments thereof, which fragments have at least 15 bases, preferably at least 30 bases, more preferably at least 50 bases and most preferably fragments having up to at least 150 bases or



greater, which fragments are at least 90% identical, preferably at least 95% identical and most preferably at least 97% identical to any portion of a polynucleotide of the present invention.

The deposit(s) referred to herein will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for purposes of Patent Procedure. These deposits are provided merely as convenience to those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained in the deposited materials, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with any description of sequences herein. A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

The present invention further relates to polypeptides which have the deduced amino acid sequences of Figures 1A-C, 3A, and 3B (SEQ ID NOS:2 and 4, respectively) as well as fragments, analogs and derivatives of such polypeptides.

The terms "fragment," "derivative," and "analog" when referring to (a) the polypeptide of Figures 1A-1C (SEQ ID NO:2) or that encoded by the deposited cDNA, or (b) the polypeptide of Figures 3A and 3B (SEQ ID NO:4), means a polypeptide which either retains substantially the same biological function or activity as such polypeptide, *i.e.*, functions as a G-protein coupled receptor, or retains the ability to bind the ligand or the receptor even though the polypeptide does not function as a G-protein coupled receptor, for example, a soluble form of the receptor.

The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a synthetic polypeptide, preferably a recombinant polypeptide.

The fragment, derivative or analog of either (a) the polypeptide of Figures 1A-1C (SEQ ID NO:2) or that encoded by the deposited cDNA, (b) the polypeptide of Figures 3A and 3B (SEQ ID NO:4) may be (I) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and

such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are f used to the mature polypeptide, or (v) one in which a fragment of the polypeptide is soluble, *i.e.*, not membrane bound, yet still binds ligands to the membrane bound receptor. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (*e.g.*, the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The polypeptides of the present invention include the polypeptides of SEQ ID NOS:2 and 4 (in particular the respective mature polypeptides) as well as polypeptides which have at least 70% similarity (preferably at least a 70% identity) to either the polypeptide of SEQ ID NO:2 or the polypeptide of SEQ ID NO:4 and more preferably at least a 90% similarity (more preferably at least a 90% identity) to the polypeptide of SEQ ID NO:2 or of SEQ ID NO:4 and still more preferably at least a 95% similarity (still more preferably a 90% identity) to the polypeptide of SEQ ID NO:2 or of SEQ ID NO:4 and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide.

Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to a method for identifying and/or isolating cells, tissues, or classes of cells or tissues, by utilizing probes of the polynucleotides that encode the EBI-2 G-protein coupled receptor polypeptide or by utilizing an antibody specific for the EBI-2 G-protein coupled receptor, for example. Since the EBI-2 G-protein coupled receptor polypeptides according to the invention occur in vein endothelial cells, neutrophil leukocyte cells, and corpus colosum cells, the above probes or antibodies, for example, may be utilized to identify and/or isolate such cells, tissues or classes of cells or tissues.

The present invention further relates to a method for identifying and/or isolating cells, tissues, or classes of cells or tissues, by utilizing probes of the polynucleotides that encode the EDG-1-like G-protein coupled receptor polypeptide or by utilizing an antibody specific for the EDG-1-like G-protein coupled receptor polypeptide, for example. Since the EDG-1-like G-protein coupled receptor polypeptides according to the invention occur in leukocyte, tonsil, placenta, thymus, lung, and pancreas tissue, the above probes or antibodies, for example, may be utilized to identify and/or isolate such cells, tissues or classes of cells or tissues.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a

cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the G-protein coupled receptor genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, *e.g.*, derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli* *lac* or *trp*, the phage lambda P<sub>L</sub> promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed

host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, *Streptomyces*, and *Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila S2* and *Spodoptera Sf9*; animal cells such as CHO, Cos or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript,  $\phi$ X174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVY3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are PKK232-8 and PCM7. Particular named bacterial promoters include *lacI*, *lacZ*, T3, T7, gpt, lambda P<sub>R</sub>, P<sub>L</sub> and *trp*. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late

SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation. (Davis, L., *et al.*, *Basic Methods in Molecular Biology*, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, Second Edition*, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples including the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons

encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK),  $\alpha$ -factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and pGEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well know to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell* 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The G-protein coupled receptor polypeptides can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The polypeptides of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. Polypeptides of the invention may also include an initial methionine amino acid residue.



The G-protein coupled receptor of the present invention may be employed in a process for screening for antagonists and/or agonists for the receptor.

In general, such screening procedures involve providing appropriate cells which express the receptor on the surface thereof. In particular, a polynucleotide encoding the receptor of the present invention is employed to transfect cells to thereby express the G-protein coupled receptor. Such transfection may be accomplished by procedures as hereinabove described.

One such screening procedure involves the use of the melanophores which are transfected to express the G-protein coupled receptor of the present invention. Such a screening technique is described in PCT WO 92/01810 published February 6, 1992.

Thus, for example, such assay may be employed for screening for a receptor antagonist by contacting the melanophore cells which encode the G-protein coupled receptor with both the receptor ligand and a compound to be screened. Inhibition of the signal generated by the ligand indicates that a compound is a potential antagonist for the receptor, *i.e.*, inhibits activation of the receptor.

The screen may be employed for determining an agonist by contacting such cells with compounds to be screened and determining whether such compound generates a signal, *i.e.*, activates the receptor.

Other screening techniques include the use of cells which express the G-protein coupled receptor (for example, transfected CHO cells) in a system which measures extracellular pH changes caused by receptor activation, for example, as described in Parce *et al.*, *Science* 246:243-247 (1989). For example, potential agonists or antagonists may be contacted with a cell which expresses the G-protein coupled receptor and a second messenger response, *e.g.*, signal transduction or pH changes, may be measured to determine whether the potential agonist or antagonist is effective.

Another such screening technique involves introducing RNA encoding the G-protein coupled receptor into *Xenopus* oocytes to transiently express the receptor. The receptor oocyte may then be contacted in the case of antagonist

screening with the receptor ligand and a compound to be screened, followed by detection of inhibition of a calcium signal.

Another screening technique involves expressing the G-protein coupled receptor in which the receptor is linked to a phospholipase C or D. As representative examples of such cells, there may be mentioned endothelial cells, smooth muscle cells, embryonic kidney cells, etc. The screening for an antagonist or agonist may be accomplished as hereinabove described by detecting activation of the receptor or inhibition of activation of the receptor from the phospholipase second signal.

Another method involves screening for antagonists by determining inhibition of binding of labeled ligand to cells which have the receptor on the surface thereof. Such a method involves transfecting a eukaryotic cell with DNA encoding the G-protein coupled receptor such that the cell expresses the receptor on its surface and contacting the cell with a potential antagonist in the presence of a labeled form of a known ligand. The ligand can be labeled, *e.g.*, by radioactivity. The amount of labeled ligand bound to the receptors is measured, *e.g.*, by measuring radioactivity of the receptors. If the potential antagonist binds to the receptor as determined by a reduction of labeled ligand which binds to the receptors, the binding of labeled ligand to the receptor is inhibited.

The present invention also provides a method for determining whether a ligand not known to be capable of binding to a G-protein coupled receptor can bind to such receptor which comprises contacting a mammalian cell which expresses a G-protein coupled receptor with the ligand under conditions permitting binding of ligands to the G-protein coupled receptor, detecting the presence of a ligand which binds to the receptor and thereby determining whether the ligand binds to the G-protein coupled receptor. The systems hereinabove described for determining agonists and/or antagonists may also be employed for determining ligands which bind to the receptor.

In general, antagonists for G-protein coupled receptors which are determined by screening procedures may be employed for a variety of therapeutic purposes. For example, such antagonists have been employed for treatment of hypertension, angina pectoris, myocardial infarction, ulcers, asthma, allergies,

psychoses, depression, migraine, vomiting, stroke, eating disorders, migraine headaches, cancer and benign prostatic hypertrophy.

Agonists for G-protein coupled receptors are also useful for therapeutic purposes, such as the treatment of asthma, Parkinson's disease, acute heart failure, hypotension, urinary retention, and osteoporosis.

Examples of G-protein coupled receptor antagonists include an antibody, or in some cases an oligonucleotide, which binds to the G-protein coupled receptor but does not elicit a second messenger response such that the activity of the G-protein coupled receptor is prevented. Antibodies include anti-idiotypic antibodies which recognize unique determinants generally associated with the antigen-binding site of an antibody. Potential antagonists also include proteins which are closely related to the ligand of the G-protein coupled receptor, *i.e.*, a fragment of the ligand, which has lost biological function and when binding to the G-protein coupled receptor, elicit no response.

A potential antagonist also includes an antisense construct prepared through the use of antisense technology. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide to DNA or RNA. For example, the 5' coding portion of the polynucleotide sequence, which encodes for the mature polypeptides of the present invention, is used to design an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription (triple-helix *see* Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251: 1360 (1991)), thereby preventing transcription and the production of G-protein coupled receptor. The antisense RNA oligonucleotide hybridizes to the mRNA *in vivo* and blocks translation of the mRNA molecule into the G-protein coupled receptor (antisense - Okano, *J. Neurochem.* 56:560 (1991) ; Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed *in vivo* to inhibit production of G-protein coupled receptor.

Another potential antagonist is a small molecule which binds to the G-protein coupled receptor, making it inaccessible to ligands such that normal biological activity is prevented. Examples of small molecules include but are not limited to small peptides or peptide-like molecules.

Potential antagonists also include a soluble form of a G-protein coupled receptor, *e.g.*, a fragment of the receptor, which binds to the ligand and prevents the ligand from interacting with membrane bound G-protein coupled receptors.

The G-protein coupled receptor and antagonists or agonists may be employed in combination with a suitable pharmaceutical carrier. Such compositions comprise a therapeutically effective amount of the polypeptide, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation should suit the mode of administration.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the pharmaceutical compositions may be employed in conjunction with other therapeutic compounds.

The pharmaceutical compositions may be administered in a convenient manner such as by the topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, the pharmaceutical compositions will be administered in an amount of at least about 10  $\mu\text{g/kg}$  body weight and in most cases they will be administered in an amount not in excess of about 8 mg/kg body weight per day. In most cases, the dosage is from about 10  $\mu\text{g/kg}$  to about 1 mg/kg body weight daily, taking into account the routes of administration, symptoms, etc.

The G-protein coupled receptor polypeptides and antagonists or agonists which are polypeptides, may be employed in accordance with the present invention by expression of such polypeptides *in vivo*, which is often referred to as "gene therapy."

Thus, for example, cells from a patient may be engineered with a polynucleotide (DNA or RNA) encoding a polypeptide *ex vivo*, with the engineered cells then being provided to a patient to be treated with the polypeptide. Such methods are well known in the art. For example, cells may be engineered by procedures known in the art by use of a retroviral particle containing RNA encoding a polypeptide of the present invention.

Similarly, cells may be engineered *in vivo* for expression of a polypeptide *in vivo* by, for example, procedures known in the art. As known in the art, a producer cell for producing a retroviral particle containing RNA encoding the polypeptide of the present invention may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in vivo*. These and other methods for administering a polypeptide of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention. For example, the expression vehicle for engineering cells may be other than a retrovirus, for example, an adenovirus which may be used to engineer cells *in vivo* after combination with a suitable delivery vehicle.

Retroviruses from which the retroviral plasmid vectors hereinabove mentioned may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen necrosis virus, retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, gibbon ape leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumor virus. In one embodiment, the retroviral plasmid vector is derived from Moloney Murine Leukemia Virus.

The vector includes one or more promoters. Suitable promoters which may be employed include, but are not limited to, the retroviral LTR; the SV40 promoter; and the human cytomegalovirus (CMV) promoter described in Miller, *et al.*, *BioTechniques* 7:980-990 (1989), or any other promoter (*e.g.*, cellular promoters such as eukaryotic cellular promoters including, but not limited to, the

histone, pol III, and  $\beta$ -actin promoters). Other viral promoters which may be employed include, but are not limited to, adenovirus promoters, thymidine kinase (TK) promoters, and B19 parvovirus promoters. The selection of a suitable promoter will be apparent to those skilled in the art from the teachings contained herein.

The nucleic acid sequence encoding the polypeptide of the present invention is under the control of a suitable promoter. Suitable promoters which may be employed include, but are not limited to, adenoviral promoters, such as the adenoviral major late promoter; or heterologous promoters, such as the cytomegalovirus (CMV) promoter; the respiratory syncytial virus (RSV) promoter; inducible promoters, such as the MMT promoter, the metallothionein promoter; heat shock promoters; the albumin promoter; the ApoAI promoter; human globin promoters; viral thymidine kinase promoters, such as the Herpes Simplex thymidine kinase promoter; retroviral LTRs (including the modified retroviral LTRs hereinabove described); the  $\beta$ -actin promoter; and human growth hormone promoters. The promoter also may be the native promoter which controls the gene encoding the polypeptide.

The retroviral plasmid vector is employed to transduce packaging cell lines to form producer cell lines. Examples of packaging cells which may be transfected include, but are not limited to, the PE501, PA317,  $\Psi$ -2,  $\Psi$ -AM, PA12, T19-14X, VT-19-17-H2,  $\Psi$ CRE,  $\Psi$ CRIP, GP+E-86, GP+envAm12, and DAN cell lines as described in Miller, *Human Gene Therapy* 1:5-14 (1990), which is incorporated herein by reference in its entirety. The vector may transduce the packaging cells through any means known in the art. Such means include, but are not limited to, electroporation, the use of liposomes, and  $\text{CaPO}_4$  precipitation. In one alternative, the retroviral plasmid vector may be encapsulated into a liposome, or coupled to a lipid, and then administered to a host.

The producer cell line generates infectious retroviral vector particles which include the nucleic acid sequence(s) encoding the polypeptides. Such retroviral vector particles then may be employed, to transduce eukaryotic cells, either *in vitro* or *in vivo*. The transduced eukaryotic cells will express the nucleic acid sequence(s) encoding the polypeptide. Eukaryotic cells which may be transduced

include, but are not limited to, embryonic stem cells, embryonic carcinoma cells, as well as hematopoietic stem cells, hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial cells.

G-protein coupled receptors are ubiquitous in the mammalian host and are responsible for many biological functions, including many pathologies. Accordingly, it is desirous to find compounds which stimulate a G-protein coupled receptor and compounds which antagonize a G-protein coupled receptor.

This invention further provides a method of identifying compounds which specifically interact with, and bind to, the human G-protein coupled receptors on the surface of a cell which comprises contacting a mammalian cell comprising an isolated DNA molecule encoding the G-protein coupled receptor with a plurality of compounds, determining those which bind to the mammalian cell, and thereby identifying compounds which specifically interact with and bind to a human G-protein coupled receptor of the present invention.

This invention also provides a method of detecting expression of the G-protein coupled receptor on the surface of a cell by detecting the presence of mRNA coding for a G-protein coupled receptor which comprises obtaining total mRNA from the cell and contacting the mRNA so obtained with a nucleic acid probe comprising a nucleic acid molecule of at least 15 nucleotides capable of specifically hybridizing with a sequence included within the sequence of a nucleic acid molecule encoding a human G-protein coupled receptor under hybridizing conditions, detecting the presence of mRNA hybridized to the probe, and thereby detecting the expression of the G-protein coupled receptor by the cell.

This invention is also related to the use of the G-protein coupled receptor gene as part of a diagnostic assay for detecting diseases or susceptibility to diseases related to the presence of mutated G-protein coupled receptor genes. Such diseases are related to cell transformation, such as tumors and cancers.

Individuals carrying mutations in the human G-protein coupled receptor gene may be detected at the DNA level by a variety of techniques. Nucleic acids for diagnosis may be obtained from a patient's cells, such as from blood, urine, saliva, tissue biopsy, and autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki *et al.*,

*Nature* 324:163-166 (1986)) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid encoding the G-protein coupled receptor protein can be used to identify and analyze G-protein coupled receptor mutations. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled G-protein coupled receptor RNA or alternatively, radiolabeled G-protein coupled receptor antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels with or without denaturing agents. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (*see, e.g., Myers et al., Science* 230:1242 (1985)).

Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (*e.g., Cotton et al., Proc. Natl. Acad. Sci. USA* 85:4397-4401 (1985)).

Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, (*e.g., Restriction Fragment Length Polymorphisms (RFLP)*) and Southern blotting of genomic DNA.

In addition to more conventional gel-electrophoresis and DNA sequencing, mutations can also be detected by *in situ* analysis.

The present invention also relates to a diagnostic assay for detecting altered levels of soluble forms of the receptor polypeptides of the present invention in various tissues. Assays used to detect levels of the soluble receptor polypeptides in a sample derived from a host are well known to those of skill in



the art and include radioimmunoassays, competitive-binding assays, Western blot analysis and preferably as ELISA assay.

An ELISA assay initially comprises preparing an antibody specific to antigens of the G-protein coupled receptor polypeptides, preferably a monoclonal antibody. In addition a reporter antibody is prepared against the monoclonal antibody. To the reporter antibody is attached a detectable reagent such as radioactivity, fluorescence or in this example a horseradish peroxidase enzyme. A sample is now removed from a host and incubated on a solid support, *e.g.*, a polystyrene dish, that binds the proteins in the sample. Any free protein binding sites on the dish are then covered by incubating with a non-specific protein such as bovine serum albumin. Next the monoclonal antibody is incubated in the dish during which time the monoclonal antibodies attach to any G-protein coupled receptor proteins attached to the polystyrene dish. All unbound monoclonal antibody is washed out with buffer. The reporter antibody linked to horseradish peroxidase is now placed in the dish resulting in binding of the reporter antibody to any monoclonal antibody bound to G-protein receptor proteins. Unattached reporter antibody is then washed out. Peroxidase substrates are then added to the dish and the amount of color developed in a given time period is a measurement of the amount of G-protein coupled receptor proteins present in a given volume of patient sample when compared against a standard curve.

The sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the cDNA. Computer analysis of the 3' untranslated region is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process.

These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 50 or 60 bases. For a review of this technique, see Verma *et al.*, *Human Chromosomes: a Manual of Basic Techniques*, Pergamon Press, New York (1988).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in McKusick, V., *Mendelian Inheritance in Man* (available on line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 (one) megabase mapping resolution and one gene per 20 kb).

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, *Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* 4:72 (1983)), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, *et al.*, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96 (1985)).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic polypeptide products of this invention.

The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 µg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 µl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. *et al.*, *Nucleic Acids Res.* 8:4057 (1980).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments Sambrook *et al.*, *ibid.* Unless

otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units to T4 DNA ligase ("ligase") per 0.5 µg of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in the method of Graham, F. and Van der Eb, A., *Virology* 52:456-457 (1973).

### ***Example 1***

#### ***Bacterial Expression and Purification of EBI-2***

The DNA sequence encoding EBI-2, ATCC #209003, is initially amplified using PCR oligonucleotide primers corresponding to the 5' sequences of the processed EBI-2 protein (minus the signal peptide sequence) and the vector sequences 3' to the EBI-2 gene. Additional nucleotides corresponding to EBI-2 were added to the 5' and 3' sequences respectively. The 5' oligonucleotide primer has the sequence 5' CCGAGGATCCATGCAAGCCGTCGACAAT 3' (SEQ ID NO:5) contains a BamHI restriction enzyme site followed by 18 nucleotides of the EBI-2 coding sequence starting from the presumed terminal amino acid of the processed protein codon. The 3' sequence 5' CCGAGGATCCTTACATTGGAGTCTCTTC 3' (SEQ ID NO:6) contains complementary sequences to BamHI site and is followed by 18 nucleotides of EBI-2. The restriction enzyme sites correspond to the restriction enzyme sites on the bacterial expression vector pQE-60 (Qiagen, Inc., Chatsworth, CA, 91311). pQE-60 encodes antibiotic resistance (Amp<sup>r</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites. pQE-60 was then digested with BamHI. The amplified sequences were ligated into pQE-60 and were inserted in frame with the sequence encoding for the histidine tag and the RBS. The ligation mixture was then used to transform *E. coli* strain M15/rep 4 (Qiagen, Inc.) by the procedure described in Sambrook, J. *et al.*, *ibid.* M15/rep4 contains multiple copies of the plasmid pREP4, which expresses the *lacI* repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants are identified by their ability to grow on LB plates, and ampicillin/kanamycin resistant colonies were selected. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the

desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 µg/ml) and Kan (25 µg/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an optical density 600 (O.D.<sub>600</sub>) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalactopyranoside") was then added to a final concentration of 1 mM. IPTG induces by inactivating the *lacI* repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra 3 to 4 hours. Cells were then harvested by centrifugation. The cell pellet was solubilized in the chaotropic agent 6 Molar Guanidine HCl. After clarification, solubilized EBI-2 was purified from this solution by chromatography on a Nickel-Chelate column under conditions that allow for tight binding by proteins containing the 6-His tag (Hochuli, E., *et al.*, *J. Chromatography* 411:177-184 (1984)). EBI-2 (95%) pure was eluted from the column in 6 molar guanidine HCl pH 5.0 and for the purpose of renaturation adjusted to 3 molar guanidine HCl, 100mM sodium phosphate, 10 mM glutathione (reduced) and 2 mM glutathione (oxidized). After incubation in this solution for 12 hours the protein was dialyzed to 10 mM sodium phosphate.

## Example 2

### *Cloning and Expression of EBI-2 Using the Baculovirus Expression System*

The DNA sequence encoding the full length EBI-2 protein, ATCC #209003, was amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the gene:

The 5' primer has the sequence 5' CCGAGGATCCGCC ATC**ATG**CAAGCCGTCGACAAT 3' (SEQ ID NO:7) and contains a BamHI restriction enzyme site (in bold) followed by 6 nucleotides resembling an efficient signal for the initiation of translation in eukaryotic cells (Kozak, M., *J. Mol. Biol.*, 196:947-950 (1987)) which is just behind the first 18 nucleotides of the EBI-2 gene (the initiation codon for translation "ATG" is underlined).

The 3' primer has the sequence 5' CCGAGGATCCTTACA-TTGGAGTCTCTTC 3' (SEQ ID NO: 8) and contains the cleavage site for the restriction endonuclease BamHI and 18 nucleotides complementary to the 3' translated sequence of the extracellular part of EBI-2. The amplified sequences

were isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment was then digested with the endonucleases BamHI, and purified again on a 1% agarose gel. This fragment is designated F2.

The vector pA2 (modification of pVL941 vector, discussed below) is used for the expression of the EBI-2 protein using the baculovirus expression system (for review see: Summers, M.D. and Smith, G.E., *A manual of methods for baculovirus vectors and insect cell culture procedures*, Texas Agricultural Experimental Station Bulletin No. 1555 (1987)). This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by the recognition sites for the restriction endonucleases BamHI. The polyadenylation site of the simian virus SV40 is used for efficient polyadenylation. For an easy selection of recombinant virus the beta-galactosidase gene from *E.coli* is inserted in the same orientation as the polyhedrin promoter followed by the polyadenylation signal of the polyhedrin gene. The polyhedrin sequences are flanked at both sides by viral sequences for the cell-mediated homologous recombination of co-transfected wild-type viral DNA. Many other baculovirus vectors could be used in place of pA2 such as pRG1 and pA2 -GP in which case the 5' primer are changed accordingly, and pAc373, pVL941 and pAcIM1 (Luckow, V.A. and Summers, M.D., *Virology* 170:31-39).

The plasmid was digested with the restriction enzyme BamHI and then dephosphorylated using calf intestinal phosphatase by procedures known in the art. The DNA was then isolated from a 1% agarose gel using the commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.). This vector DNA is designated V2.

Fragment F2 and the dephosphorylated plasmid V2 were ligated with T4 DNA ligase. *E.coli* HB101 cells were then transformed and bacteria identified that contained the plasmid (pBacEBI-2) with the EBI-2 gene using the enzyme BamHI. The sequence of the cloned fragment was confirmed by DNA sequencing.

5 µg of the plasmid pBacEBI-2 was co-transfected with 1.0 µg of a commercially available linearized baculovirus ("BaculoGold™ baculovirus DNA",

Pharmingen, San Diego, CA.) using the lipofection method (Felgner *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:7413-7417 (1987)).

1 µg of BaculoGold™ virus DNA and 5 µg of the plasmid pBacEBI-2 were mixed in a sterile well of a microtiter plate containing 50 µl of serum free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards 10 µl Lipofectin plus 90 µl Grace's medium were added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture was added drop-wise to the Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate was rocked back and forth to mix the newly added solution. The plate was then incubated for 5 hours at 27°C. After 5 hours the transfection solution was removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum (FBS) was added. The plate was put back into an incubator and cultivation continued at 27°C for four days.

After four days the supernatant was collected and a plaque assay performed similar as described by Summers and Smith (*supra*). As a modification an agarose gel with "Bluo-gal" (Life Technologies Inc., Gaithersburg) was used which allows an easy isolation of blue stained plaques. (A detailed description of a "plaque assay" can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10).

Four days after the serial dilution, the virus was added to the cells and blue stained plaques were picked with the tip of an Eppendorf pipette. The agar containing the recombinant viruses was then resuspended in an Eppendorf tube containing 200 µl of Grace's medium. The agar was removed by a brief centrifugation and the supernatant containing the recombinant baculovirus was used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes were harvested and then stored at 4°C.

Sf9 cells were grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells were infected with the recombinant baculovirus V-EBI-2 at a multiplicity of infection (MOI) of 2. Six hours later the medium was removed and replaced with SF900 II medium minus methionine and cysteine (Life Technologies Inc., Gaithersburg). 42 hours later 5 µCi of <sup>35</sup>S-methionine and 5



$\mu\text{Ci } ^{35}\text{S}$  cysteine (Amersham) were added. The cells were further incubated for 16 hours before they were harvested by centrifugation and the labeled proteins visualized by SDS-PAGE and autoradiography.

### ***Example 3***

#### ***Expression of Recombinant EBI-2 in COS cells***

The expression of plasmid, EBI-2 HA is derived from a vector pcDNAI/Amp (Invitrogen) containing: 1) SV40 origin of replication, 2) ampicillin resistance gene, 3) *E. coli* replication origin, 4) CMV promoter followed by a polylinker region, an SV40 intron, and polyadenylation site. A DNA fragment encoding the entire EBI-2 precursor and a HA tag fused in frame to its 3' end was cloned into the polylinker region of the vector, therefore, the recombinant protein expression is directed under the CMV promoter. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein as previously described (Wilson, I., *et al.*, *Cell* 37:767 (1984)). The fusion of HA tag to the target protein allows easy detection of the recombinant protein with an antibody that recognizes the HA epitope.

The plasmid construction strategy is described as follows:

The DNA sequence encoding EBI-2, ATCC #209003, was constructed by PCR using two primers: the 5' primer 5' CCGAGGATCCGCCA-TCATGCAAGCCGTCGACAAT 3' (SEQ ID NO:9) contains a BamHI site followed by EBI-2 coding sequence starting from the initiation codon; the 3' sequence 5' CCGATCTAGATTAATCCCATACGACGTCCTCCAGACTACGC-TCATTGGAGTCTCTTC 3' (SEQ ID NO:10) contains complementary sequences to XbaI site, translation stop codon, HA tag and EBI-2 coding sequence (not including the stop codon). Therefore, the PCR product contains a BamHI site, EBI-2 coding sequence followed by HA tag fused in frame, a translation termination stop codon next to the HA tag, and an XbaI site. The PCR amplified DNA fragment and the vector, pcDNAI/Amp, were digested with BamHI and XbaI restriction enzyme and ligated. The ligation mixture was transformed into *E. coli* strain SURE (available from Stratagene Cloning Systems, 11099 North Torrey Pines Road, La Jolla, CA 92037) the transformed culture was plated on

ampicillin media plates and resistant colonies were selected. Plasmid DNA was isolated from transformants and examined by restriction analysis for the presence of the correct fragment. For expression of the recombinant EBI-2, COS cells were transfected with the expression vector by DEAE-DEXTRAN method (Sambrook, J., *et al.*, *ibid.*). The expression of the EBI-2 HA protein was detected by radiolabelling and immunoprecipitation method (Harlow, E., and Lane, D., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, (1988)). Cells were labelled for 8 hours with <sup>35</sup>S-cysteine two days post transfection. Culture media was then collected and cells were lysed with detergent (RIPA buffer (150 mM NaCl, 0.1% SDS, 1% NP-40, 0.5% DOC, 50mM Tris, pH 7.5) (Wilson, I. *et al.*, *Id.* 37:767 (1984)). Both cell lysate and culture media were precipitated with an HA specific monoclonal antibody. Proteins precipitated were analyzed on 15% SDS-PAGE gels.

### ***Example 4***

#### ***Expression via Gene Therapy***

Fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (*e.g.*, Ham's F12 media, with 10% FBS, penicillin and streptomycin, is added). This is then incubated at 37°C for approximately one week. At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pmV-7 (Kirschmeier, P.T., *et al.*, *DNA* 7:219-225 (1988)) flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention is amplified using PCR primers which correspond to the 5' and 3' end sequences respectively. The 5' primer containing an EcoRI site and the 3' primer further includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is used to transform bacteria HB101, which are then plated onto agar-containing kanamycin for the purpose of confirming that the vector had the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells are transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his.

The engineered fibroblasts are then injected into the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads. The fibroblasts now produce the protein product.

### ***Example 5***

#### ***Bacterial Expression and Purification of EDG-1-Like Polypeptide***

The DNA sequence encoding EDG-1-like polypeptide, ATCC #209004, is initially amplified using PCR oligonucleotide primers corresponding to the 5' sequences of the processed EDG-1-like polypeptide protein (minus the signal peptide sequence) and the vector sequences 3' to the EDG-1-like polypeptide gene. Additional nucleotides corresponding to EDG-1-like sequence were added to the 5' and 3' sequences respectively. The 5' oligonucleotide primer has the sequence 5' CCGAGGATCCATGAACGCCACGGGGACC 3' (SEQ ID NO:11) contains a BamHI restriction enzyme site followed by 18 nucleotides of the EDG-1-like polypeptide coding sequence starting from the presumed terminal amino acid of the processed protein codon. The 3' sequence 5' CCGAGGATCCTCAGATGCTCCGCACGCT 3' (SEQ ID NO:12) contains complementary sequences to BamHI site and is followed by 18 nucleotides of EDG-1-like polypeptide. The restriction enzyme sites correspond to the restriction enzyme sites on the bacterial expression vector pQE-60 (Qiagen, Inc., Chatsworth, CA, 91311). pQE-60 encodes antibiotic resistance (Amp<sup>r</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites. pQE-60 was then digested with BamHI. The amplified sequences were ligated into pQE-60 and were inserted in frame with the sequence encoding for the histidine tag and the RBS. The ligation mixture was then used to transform *E. coli* strain M15/rep 4 (Qiagen, Inc.) by the procedure described in Sambrook, J., *et al.*, *ibid*. M15/rep4 contains multiple copies of the plasmid pREP4, which expresses the *lacI* repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies were selected. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 µg/ml) and Kan (25 µg/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an optical density 600 (O.D.<sub>600</sub>) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalactopyranoside") was then added to a final concentration of 1 mM. IPTG induces by inactivating the *lacI* repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra

The vector pA2 (modification of pVL941 vector, discussed below) is used for the expression of the EDG-1-like polypeptide protein using the baculovirus expression system (for review see: Summers, M.D. and Smith, G.E., *A manual of methods for baculovirus vectors and insect cell culture procedures*, Texas Agricultural Experimental Station Bulletin No. 1555 (1987)). This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by the recognition sites for the restriction endonucleases BamHI. The polyadenylation site of the simian virus SV40 is used for efficient polyadenylation. For an easy selection of recombinant virus the beta-galactosidase gene from *E.coli* is inserted in the same orientation as the polyhedrin promoter followed by the polyadenylation signal of the polyhedrin gene. The polyhedrin sequences are flanked at both sides by viral sequences for the cell-mediated homologous recombination of co-transfected wild-type viral DNA. Many other baculovirus vectors could be used in place of pA2 such as pRG1 and pA2-GP in which case the 5' primer are changed accordingly, and pAc373, pVL941 and pAcIM1 (Luckow, V.A. and Summers, M.D., *Virology* 170:31-39).

The plasmid was digested with the restriction enzyme BamHI and then dephosphorylated using calf intestinal phosphatase by procedures known in the art. The DNA was then isolated from a 1% agarose gel using the commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.) This vector DNA is designated V2.

Fragment F2 and the dephosphorylated plasmid V2 were ligated with T4 DNA ligase. *E.coli* HB101 cells were then transformed and bacteria identified that contained the plasmid (pBacEDG-1-like polypeptide) with the EDG-1-like polypeptide gene using the enzyme BamHI. The sequence of the cloned fragment was confirmed by DNA sequencing.

5 µg of the plasmid pBacEDG-1-like polypeptide was co-transfected with 1.0 µg of a commercially available linearize baculovirus ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA.) using the lipofection method (Felgner *et al.*, *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987)).

1 µg of BaculoGold™ virus DNA and 5 µg of the plasmid pBacEDG-1-like polypeptide were mixed in a sterile well of a microtiter plate containing 50 µl of serum free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards 10 µl Lipofectin plus 90 µl Grace's medium were added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture was added drop-wise to the Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate was rocked back and forth to mix the newly added solution. The plate was then incubated for 5 hours at 27°C. After 5 hours the transfection solution was removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum was added. The plate was put back into an incubator and cultivation continued at 27°C for four days.

After four days the supernatant was collected and a plaque assay performed similar as described by Summers and Smith (*supra*). As a modification an agarose gel with "Bluo-gal" (Life Technologies Inc., Gaithersburg) was used which allows an easy isolation of blue stained plaques. (A detailed description of a "plaque assay" can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10).

Four days after the serial dilution, the virus was added to the cells and blue stained plaques were picked with the tip of an Eppendorf pipette. The agar containing the recombinant viruses was then resuspended in an Eppendorf tube containing 200 µl of Grace's medium. The agar was removed by a brief centrifugation and the supernatant containing the recombinant baculovirus was used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes were harvested and then stored at 4°C.

Sf9 cells were grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells were infected with the recombinant baculovirus V-EDG-1-like polypeptide at a multiplicity of infection (MOI) of 2. Six hours later the medium was removed and replaced with SF900 II medium minus methionine and cysteine (Life Technologies Inc., Gaithersburg). 42 hours later 5 µCi of <sup>35</sup>S-methionine and 5 µCi <sup>35</sup>S cysteine (Amersham) were added. The cells were

further incubated for 16 hours before they were harvested by centrifugation and the labelled proteins visualized by SDS-PAGE and autoradiography.

### ***Example 7***

#### ***Expression of Recombinant EDG-1-like Polypeptide in COS Cells***

The expression of plasmid, EDG-1-like polypeptide HA is derived from a vector pcDNA1/Amp (Invitrogen) containing: 1) SV40 origin of replication, 2) ampicillin resistance gene, 3) *E. coli* replication origin, 4) CMV promoter followed by a polylinker region, an SV40 intron and polyadenylation site. A DNA fragment encoding the entire EDG-1-like polypeptide precursor and a HA tag fused in frame to its 3' end was cloned into the polylinker region of the vector, therefore, the recombinant protein expression is directed under the CMV promoter. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein as previously described. The fusion of HA tag to the target protein allows easy detection of the recombinant protein with an antibody that recognizes the HA epitope.

The plasmid construction strategy is described as follows:

The DNA sequence encoding EDG-1-like polypeptide, ATCC #209004, was constructed by PCR using two primers: the 5' primer 5' CCGAGGATC-CGCCATCATGAACGCCACGGGGACC 3' (SEQ ID NO:15) contains a BamHI site followed by EDG-1-like polypeptide coding sequence starting from the initiation codon; the 3' sequence 5' CCGATCTAGATCAATCCCATACG-ACGTCCCAGACTACGCTGATGCTCCGCACGCT 3' (SEQ ID NO:16) contains complementary sequences to XbaI site, translation stop codon, HA tag and EDG-1-like polypeptide coding sequence (not including the stop codon). Therefore, the PCR product contains a BamHI site, EDG-1-like polypeptide coding sequence followed by HA tag fused in frame, a translation termination stop codon next to the HA tag, and an XbaI site. The PCR amplified DNA fragment and the vector, pcDNA1/Amp, were digested with BamHI and XbaI restriction enzyme and ligated. The ligation mixture was transformed into *E. coli* strain SURE (available from Stratagene Cloning Systems, 11099 North Torrey Pines Road, La Jolla, CA 92037) the transformed culture was plated on ampicillin media

plates and resistant colonies were selected. Plasmid DNA was isolated from transformants and examined by restriction analysis for the presence of the correct fragment. For expression of the recombinant EDG-I-like polypeptide COS cells were transfected with the expression vector by DEAE-DEXTRAN method. The expression of the EDG-I-like polypeptide HA protein was detected by radiolabelling and immunoprecipitation method. Cells were labelled for 8 hours with  $^{35}\text{S}$ -cysteine two days post transfection. Culture media was then collected and cells were lysed with detergent (RIPA buffer). Both cell lysate and culture media were precipitated with an HA specific monoclonal antibody. Proteins precipitated were analyzed on 15% SDS-PAGE gels.

### ***Example 8***

#### ***Expression via Gene Therapy***

Fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin, is added. This is then incubated at 37°C for approximately one week. At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger-flasks.

pMV-7 flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention is amplified using PCR primers which correspond to the 5' and 3' end sequences respectively. The 5' primer containing an EcoRI site and the 3' primer further includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and



the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is used to transform bacteria HB101, which are then plated onto agar-containing kanamycin for the purpose of confirming that the vector had the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in DMEM with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells are transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his.

The engineered fibroblasts are then injected into the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads. The fibroblasts now produce the protein product.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described.

***WHAT IS CLAIMED IS:***

1. An isolated polynucleotide comprising a polynucleotide having at least a 95% identity to a member selected from the group consisting of:

(a) a polynucleotide encoding a polypeptide comprising amino acids 2 to 342 of SEQ ID NO:2;

(b) a polynucleotide encoding a polypeptide comprising amino acids 1 to 260 of SEQ ID NO:4; and

(c) the complement of (a) or (b).

2. The isolated polynucleotide of claim 1 wherein said member is (a).

3. The isolated polynucleotide of claim 1 wherein said member is (b).

4. The isolated polynucleotide of claim 1, wherein the polynucleotide is DNA.

4. The isolated polynucleotide of claim 1, wherein the polynucleotide is RNA.

5. A method of making a recombinant vector comprising inserting the isolated polynucleotide of claim 1 into a vector, wherein said polynucleotide is DNA.

6. A recombinant vector comprising the polynucleotide of claim 1, wherein said polynucleotide is DNA.

7. A recombinant host cell comprising the polynucleotide of claim 1, wherein said polynucleotide is DNA.

8. A method for producing a polypeptide comprising expressing from the recombinant cell of claim 11 the polypeptide encoded by the polynucleotide.

9. The isolated polynucleotide of claim 1 comprising a polynucleotide, which includes nucleotides 226-1251 of SEQ ID NO:1.
10. The isolated polynucleotide of claim 1 comprising a polynucleotide, which includes nucleotides 2 to 827 of SEQ ID NO:3.
11. An isolated polynucleotide comprising a polynucleotide having at least a 95% identity to a member selected from the group consisting of:
  - (a) a polynucleotide encoding the same polypeptide encoded by the human cDNA in ATCC Deposit No. 209003;
  - (b) a polynucleotide encoding the same polypeptide encoded by the human cDNA in ATCC Deposit No. 209004; and
  - (c) the complement of (a) or (b).
12. The isolated polynucleotide of claim 17, wherein the member is (a).
13. The isolated polynucleotide of claim 17, wherein the member is (b).
14. A method of making a recombinant vector comprising inserting the isolated polynucleotide of claim 11 into a vector, wherein said polynucleotide is DNA.
15. A recombinant vector comprising the polynucleotide of claim 11, wherein said polynucleotide is DNA.
16. A recombinant host cell comprising the polynucleotide of claim 11, wherein said polynucleotide is DNA.
17. A method for producing a polypeptide comprising expressing from the recombinant cell of claim 16 the polypeptide encoded by said polynucleotide.

18. An isolated polypeptide comprising:

a mature polypeptide having an amino acid sequence encoded by a polynucleotide which is at least 95% identical to member selected from the group consisting of:

(a) a polynucleotide encoding a polypeptide comprising amino acids 2 to 342 of SEQ ID NO:2;

(b) a polynucleotide encoding a polypeptide comprising amino acids 1 to 260 of SEQ ID NO:4; and

(c) the complement of (a) or (b).

19. An antibody against the polypeptide of claim 18.

20. An antagonist against the polypeptide of claim 18.

21. A process for diagnosing a disease or a susceptibility to a disease related to an under-expression of the polypeptide of claim 18 comprising:

determining a mutation in a nucleic acid sequence encoding said polypeptide.

Two human G-Protein coupled receptor polypeptides and DNA (RNA) encoding each of such polypeptides and a procedure for producing such polypeptides by recombinant techniques is disclosed. Also disclosed are methods for utilizing such polypeptides for identifying antagonists and agonists to such polypeptides. Also disclosed are diagnostic methods for detecting a mutation in the nucleic acid sequence of each of the G-protein coupled receptors.

A273-99spec.wpd

1 GCACGAGGAACAGAACTTTCTCATGTCCAGGGTCAGATTACAAGAGCACTCAAGACTT 60

61 TACTGACGAAAACCTCAGGAATCCTCTATCACAAGAGGTTTGGCAACTAACTAAGACA 120

121 TTAAGGAAAATACCAGATGCCACTCTGCAGGCTGCAATACTACTTACTGGATAC 180

181 ATTCAAACCTCCAGAATCAACAGTTATCAGGTAACCAACAAGAAATGCAAGCCGTCGAC 240  
1 M Q A V D 5

241 AATCTCACCTCTGCGCTGGGAACACCACTGTGTGCACCAGAGACTACAAAATACCCAG 300  
6 N L T S A P G N T S L C T R D Y K I T Q 25

301 GTCCTCTTCCACTGCTCTACACTGTCCTGTTTTTTGTTGGACTTATCACAATGGCCTG 360  
26 V L F P L L Y T V L F F V G L I T N G L 45

361 GCGATGAGGATTTTCTTTCAAATCCGGAGTAATCAAACTTTATTATTTTCTTAAGAAC 420  
46 A M R I F F Q I R S K S N F I I F L K N 65

421 ACAGTCATTTCTGATCTTCTCATGATTCTGACTTTTCCATTCAAATCTTAGTGATGCC 480  
66 T V I S D L L M I L T F P F K I L S D A 85

481 AAACGGGAACAGGACCACTGAGAACTTTTGTGTGTCAGTTACCTCCGTCATATTTTAT 540  
86 K L G T G P L R T F V C Q V T S V I F Y 105

541 TTCACAATGTATATCAGTATTTTCTTCTGGGACTGATAACTATCGATCGCTACCAGAAG 600  
106 F T M Y I S I S F L G L I T I D R Y Q K 125

601 AcCACCAGGCCATTAAAACATCCAACCCCAAAATCTCTTGGGGCTAAGATTCTCTCT 660  
126 T T R P F K T S N P K N L L G A K I L S 145

661 GTTGTATCTGGGCATTCATGTTCTTACTCTCTTtGCCTAACATGATTCTGACCAACAGg 720  
146 V V I W A F M F L L S L P N M I L T N R 165

721 CAGCCGAGAGACAAGAATGTGaAGAAaTGCTCTTTCCTTAAATCAGAGTTCGGTCTAGTC 780  
166 Q P R D K N V K K C S F L K S E F G L V 185

781 TGGCATGAAATAGTAAATTACATCTGTCAAGTCATTTCTGGATTAATTTCTTAATTGTT 840  
186 W H E I V N Y I C Q V I F W I N F L I V 205

FIG.1A

841	ATTGTATGTTATACACTCATTACAAAAGAACTGTACCGGTCATACGTAAGAACGAGGGGT	900
206	I V C Y T L I T K E L Y R S Y V R T R G	225
901	GTAGGTAAAGTCCCCAGGAAAAAGGTGAACGTCAAAGTTTTTCATTATCATTGCTGTATTC	960
226	V G K V P R K K V N V K V F I I I A V F	245
961	TTTATTGTTTTGTTCCTTTCCATTTTGCCGAATTCCTTACACCCGTAGGCCAAACCGG	1020
246	F I C F V P F H F A R I P Y T L S Q T R	265
1021	GATGCTTTGACTGCACtGcTGAAAACTCTGTTCATGTGAAAGAGAGCACTCTGTGG	1080
266	D V F D C T A E N T L F Y V K E S T L W	285
1081	TTAACTTCCTTAAATGCATGCCTGGATCCGTTCACTATTTTTCTTTGCAAGTCCTTC	1140
286	L T S L N A C L D P F I Y F F L C K S F	305
1141	AGAAATTCCTTGATAAGTATGCTGAAGTGCCCCAATTCGCAACATCTCTGTCCCAGGAC	1200
306	R N S L I S M L K C P N S A T S L S Q D	325
1201	AATAGGAAAAAGAACAGGATGGTGGTGAcCCAATGAAGAGACTCCAATGTAAACAAAT	1260
326	N R K K E Q D G G D P N E E T P M *	343
1261	TAAC TAAGGAAATATTTCAATCTCTTTGTGTTCAGAACTCGTTAAAGCAAAGCGCTAAGT	1320
1321	AAAAATATTAAC TGACGAAGAAGCAACTAAGTTAATAATAATGACTCTAAAGAAACAGAA	1380
1381	GATTACAAAAGCAATTTTCATTACCTTTCCAGTATGAAAAGCTATCTTAAATATAGAA	1440
1441	AACTAATCTAACTGTAGCTGTATTAGCAGCAAAACAAACGACATCCAATTGTCATGCTG	1500
1501	CATGCAAAACTACACAGAATTCATGTTTTGgCAGAGTTTTGGCAAAATGAGTAAATCATAT	1560
1561	AATATTTACTGTAATTTTTAAAAATACATTATCGTTCCACAATTTATTTTTTCATAATCAA	1620
1621	CTAAGGAAGAACGATCAATTGGATATAATCTTCTTACCAAAAATGATAGTTAAATGTAT	1680
1681	ATATATCCTAGTCCCCTAACCaAATCCTGACCTATTGGGATACTTATAAAAAATTAAGTA	1740
1741	AGTGGGATACACAAGAATAATAACTATTAACTTTTCATTATTAGCcAAAAACCTAAGGG	1800

FIG. 1B

1801 ATTTAACTAATTGAAaCTGTATTTGATTGGACTTAATTTTTATGTTTATTTAGAAGAT 1860  
 1861 AAAGATTTAAGAAGACCTTTACAATAAAGAGAAGAAATATCGAAGTCATTAAAAAAGGA 1920  
 1921 GACTTACTTTTATGACATTCTAATACTAAAAATATAGAAATATTTCTTAATTCTAGAG 1980  
 1981 AAAGTAGTTTTACTAATTTTTTACAACCTCAATAATACCATCACTGACACTTACCTTTAT 2040  
 2041 TAATTAGCTTCTAGAAAATAGCTGCTAATTAGGTTAATGAACATTTTACCTTAGTGAAAA 2100  
 2101 AAAaTTAATTAATATGATTACAAAGTTGCACAGCATAACTACTGAGAGGAAAGTGATTG 2160  
 2161 ATCTGTTTGAATTACTTGTTTGTATTGGTGTGTATAAAATACAAATTTACATTAAACTC 2220  
 2221 TAAAtcattaaaAAAAAAAAAAAAAAAA 2247

FIG.1C



1	MQAVDNLTSAP...GNTSLCTRDKITQVLFPLLYTVLFFVGLITNGLA	46
	:  : : :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
3	IQMANNFTPPSATPQGNCDLYAHHSTARIVMPLHYSLVFIIGLVGNLLA	52
47	MRIFQIRSKSN.FIIFLKNTVISDLLMILTFPFKILSDAKLTGTPLRTF	95
	: :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
53	LVVIVQNRKKINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDA	102
96	VCQVTSVIFYFTMYISISFLGLITIDRYQKTTRPFKTSNPNKLLGAKILS	145
	:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
103	LCRITALVFYINTYAGVNFMTCLSIDRFIAVVHPLRYNKKRIEHAKGVC	152
146	VVIWAFMFLSLPNMI...LTNRQPRDKNVKKCSFLKSEGLVWHEIVNYI	193
	:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
153	IFVWILVFAQTLPLLINPMSKQEAERITCMEYPNFEETKSLPWILLGACF	202
194	CQVIFWINFLIVICYTLITKELYRSYVRTRGVGK..VPRKKVNVKVFI	241
	:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
203	IGYVL..PLIIILICYSQICCKLFRTAQNPLTEKSGVNKKALNTIILII	250
242	IAVFFICFVPFHFAPIPYTLSQLTR..DVFDCTAENTLFYVKESTLWLTSL	289
	:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
251	V.VFVLCFTPYPHVAIIQHMIKKLRFNSNFLECSQRHSFQISLHFTVCLMNF	299
290	NACLDPIFYFFLCKSFRNSLISMLKCPNSATSLSQDNRKKEQDGGDPNEE	339
	:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
300	NCCMDPIFYFFACKGYKRKVMRMLKRQVS.VSISSAVKSAPEENSREMT	348
340	TPM	342
	:	
349	TQM	351

FIG.2

1 GGACAGAGCCACCCTGCGTCGGGCCTCAGTCAGCCCCGGGGAGGCCATGAACGCCAC 60  
1 M N A T 4

61 GGGGACCCCGGTGGCCCCGAGTCCTGCCAACAGCTGGCGGCCGGCGGGCACAGCCGGCT 120  
5 G T P V A P E S C Q Q L A A G G H S R L 24

121 CATTGTTCTGCACTACAACCACTCGGGCCGGCTGGCCGGGCGCGGGGGCCGGAGGATGG 180  
25 I V L H Y N H S G R L A G R G G P E D G 44

181 CGGCCTGGGGGCCCTGCGGGGGCTGTCGGTGGCCGCCAGCTGCCTGGTGGTGTGGAGAA 240  
45 G L G A L R G L S V A A S C L V V L E N 64

241 CTTGCTGGTGTGGCGGCCATCACCAGCCACATGCGGTGCGAACGCTGGGTCTACTATTG 300  
65 L L V L A A I T S H M R S Q R W V Y Y C 84

301 CCTGGTGAACATTACGATGAGTGACCTGCTCACGGGCGCGGCCACCTGGCCAACTGCT 360  
85 L V N I T M S D L L T G A A Y L A N V L 104

361 GCTGTCGGGGGCCCGCACCTTCGGTCTGGCGCCCGCCAGTGGTTCTACGGAAGGGCCT 420  
105 L S G A R T F R L A P A Q W F L R K G L 124

421 GCTCTTCACCGCCCTGGCGCCTCCACCTTCAGCCTGCTCTTCACTGCAGGGTTGCGCTT 480  
125 L F T A L A A S T F S L L F T A G L R F 144

481 TGCCACCATGGTGCGGCCGGTGGCCGAGAGCGGGGCCACCAAGACAGCCGCGTCTACGG 540  
145 A T M V R P V A E S G A T K T S R V Y G 164

541 CTTTCATCGGCCTCTGCTGGCTGCTGGCCGCGCTGCTGGGGATGCTGCCTTTGCTGGGCTG 600  
165 F I G L C W L L A A L L G M L P L L G W 184

601 GAACTGCCTGTGCGCCTTTGACCGCTGCTCCAGCCTTCTGCCCTCTACTCCAAGCGCTA 660  
185 N C L C A F D R C S S L L P L Y S K R Y 204

661 CATCCTCTTCTGCCTGGTGATCTTCGCCGGCGTCTGGCCACCATCATGGCCTCTATGG 720  
205 I L F C L V I F A G V L A T I M G L Y G 224

721 GGCCATCTTCCGCTGGTGCAGGCCAGCGGGCAGAAGGCCCCACGCCAGCGGCCCGCGG 780  
225 A I F R L V Q A S G Q K A P R P A A R R 244

FIG.3A

781	CAAGGCCCGCCGCCTGCTGAAGACGGTGTGATGATCCTGCTGGCCCTTCTTGGTGTGCTG	840
245	K A R R L L K T V L M I L L A F L V C W	264
841	GGGACCACTCTTCGGGCTGCTGCTGGCCGACGTCCTTGGCTCCAACCTCTGGGCCAGGA	900
265	G P L F G L L L A D V F G S N L W A Q E	284
901	GTACCTGCGGGGCATGAGCTGGATCCTGGCCCTGGCCGCTCCTCAACTCGGCGGTCAACCC	960
285	Y L R G M D W I L A L A V L N S A V N P	304
961	CATCATCTACTCCTTCGCGAGCAGGGAGGTGTGCAGAGCCGTGCTCAGCTTCCTCTGCTG	1020
305	I I Y S F R S R E V C R A V L S F L C C	324
1021	CGGGTGCTCCGGCTGGGCATGCGAGGGCCCGGGAGCTGCCTGGCCCGGGCCGTGAGGC	1080
325	G C L R L G M R G P G D C L A R A V E A	344
1081	TCACTCGGAGCTTCCACCACCGACAGCTCTCTGAGGCCAAGGGACAGCTTTCGCGGCTC	1140
345	H S G A S T T D S S L R P R D S F R G S	364
1141	CCGCTCGCTCAGCTTTCGGATCGGGAGCCCTGTCCAGCATCTCCAGCGTGGGAGCAT	1200
365	R S L S F R M R E P L S S I S S V R S I	384
1201	CTGAAGTTGCAGTCTTGCGTGGATGGTGTCAACCACCGGGTGCCTGCCAGGCAGGCCCT	1260
385	*	385
1261	CCTGGGGTACAGGAAGCTGTGTGCACGCAACCTCGCCCTGTATGGGGAGCAGGGAACGGG	1320
1321	ACAGGCCCCCATGGACTTGCCCGGTGGCCTCTCGGGGCTTCTGACGCCATATGGACTTGC	1380
1381	CCATTGCCTATGGCTCACCTTGGACAAGGAGGCAACCAACCCACCTCCCCGTAGGAGCAG	1440
1441	AGAGCACCTTGGTGTGGGGGCGAGTGGGTTCCCCACAACCCGCTTCTGTGTGATTCTGG	1500
1501	GGAAGTCCCGGCCCTCTCTGGGCCTCAGTAGGGCTCCAGGCTGCAAGGGGTGGACTGT	1560
1561	GGGATGCATGCCCTGGCAACATTGAAGTTCGATCATGGTAAAAAAAAAAAAAAAAAAAAA	1620
1621	AAAAAAAAAAAAAAAAAAAA	1637

FIG. 3B



**DECLARATION FOR PATENT APPLICATION**

As a below named inventor, I declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

**HUMAN G-PROTEIN COUPLED RECEPTORS**

the specification of which [ ] is attached hereto or [x] was filed on May 7, 1997 as Application Serial No. 08/852,824 and was amended on \_\_\_\_\_ (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed. Prior Foreign Application(s):

Priority Claimed

Yes No

(Number)

(Country)

(Day/Month/Year Filed)

I hereby claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

(Application Serial No.)

(Filing Date)

(Status: patented, pending, abandoned)

I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith: John N. Bain (Reg. No. 18,651); John G. Gilfillan, III (Reg. No. 22,746); Elliot M. Olstein (Reg. No. 24,025); Raymond J. Lillie (Reg. No. 31,778); Charles J. Herron (Reg. No. 28,019); Gregory Ferraro (Reg. No. 36,134); William Squire (Reg. No. 25,378) of Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road, Roseland, New Jersey, 07068, Robert H. Benson (Reg. No. 30, 446) and Larry S. Millstein (Reg. No. 34,679) of Human Genome Sciences, Inc. 9410 Key West Avenue, Rockville, Maryland, 20878.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of first joint inventor: YI LI

Inventor's signature: [Signature]

Date: 12/17/97

Residence: 1247 Lakeside Drive, Sunnyvale, California 94086

Citizenship: China

Post Office Address: Same

Full name of second joint inventor: STEVEN M. RUBEN

Inventor's signature: [Signature]

Date: 12/22/97

Residence: 18528 Heritage Hills Drive, Olney, Maryland 20823

Citizenship: USA

Post Office Address: Same

Docket No. PF351.SKB

CB #: 325800-600

**CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI, STEWART & OLSTEIN**  
6 Becker Farm Road - Roseland, NJ 07068 - (201) 994-1700

steg/ipopen/app/pf351po/dec/c11

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Li *et al.*

Appl. No.: 08/852,824

Filed: May 7, 1997

For: **Human G-Protein Coupled  
Receptors**

Art Unit: 1646

Examiner: Basi, N.

Atty. Docket: 1488.1220000/EKS/EJH

**Declaration of Steven M. Ruben  
Under 37 C.F.R. § 1.132**

Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

1. I, Steven M. Ruben, hereby declare and state as follows:
2. I am a named inventor of the captioned application, which is assigned to Human Genome Sciences, Inc. (HGS), and I am presently employed by HGS. The work described below was done by myself, under my supervision, or as part of a collaborative research effort with other individuals at HGS.

***Human Epstein Barr Virus-Induced G-Protein Coupled Receptor-2 (EBI-2)***

3. We obtained a cDNA clone encoding a human Epstein Barr virus-induced G-protein coupled receptor-2 (EBI-2) by screening a human hippocampus cDNA library. This clone was designated HHPGS02. We determined nucleotide sequence information for the HHPGS02 clone, as described below, using sequencing methods which were routine and publicly available as of the May 7, 1997 filing date of the present application. The HHPGS02 clone that we obtained this sequence information from was deposited with the American Type Culture

Collection (ATCC) on April 28, 1997 and was given ATCC Accession No. 209003. (See Attachment A.)

4. Evidence that the human HHPGS02 cDNA was deposited at the ATCC as Accession No. 209003 is provided by comparing the ATCC Deposit Receipt (Attachment A) with the information provided in the IRIS notebook page (entitled "Sequence Worksheet") included herewith as the first page of Attachment B.<sup>1</sup> The section of the page entitled "Sequence Information" corresponds clone HHPGS02<sup>2</sup> with the "HGS Code," 405439. HGS Code 405439 represents a particular sequence entry in IRIS for cDNA clone HHPGS02. HGS Code 405439 appears as the identifier on the ATCC deposit receipt. (See Attachment A.) This indicates that the clone used to obtain the sequence information of HGS Code 405439 was deposited. In other words, even though, as explained below, SEQ ID NO:1 and SEQ ID NO:2 in the Sequence Listing of the present application as originally filed, had typographical errors due to attorney error, the human HHPGS02 cDNA clone used to obtain the original, correct sequence data was deposited at the ATCC.

5. Attachment B provides four pages of data from the IRIS electronic notebook which shows the sequencing results of the human EBI-2 cDNA clone (*i.e.*, HHPGS02). The HHPGS02 sequence was obtained using a 373 Automated DNA sequencer (Applied Biosystems, Inc.). Sequencing accuracy using this method is predicted to be greater than 97%.

6. The information obtained from the HHPGS02 sequencing run differs from the Sequence Listing currently on file in the present application at four positions. In particular, SEQ

---

<sup>1</sup>IRIS is an electronic notebook used by HGS scientists to enter and maintain sequence data.

<sup>2</sup>The "XX" designation added to the 7-character clone ID on the IRIS Notebook pages, *e.g.*, HHPGS02, merely indicates that the sequence of that clone is full-length.

ID NO:1 contains typographical errors at the following nucleotide positions: position 242, which should be A rather than T; position 266, which should be C rather than A; position 1870 (in the 3' untranslated region), where a T should be deleted, and position 2206, where an N should be deleted. These typographical errors in the nucleotide sequence result in the following errors to the encoded amino acid sequence depicted as SEQ IDNO:2: an isoleucine at position 6 should be replaced with an asparagine, and an asparagine at position 14 should be replaced with a threonine. Both of these changes are reflected in the HHPGS02 amino acid sequence data shown on the third and fourth pages of Attachment B, as well as in an amino acid alignment originally filed with the present application as Figure 2.

7. I believe that the actual nucleotide sequence of the human HHPGS02 cDNA clone is the same as that originally entered in the IRIS notebook.

8. I am of the opinion that the correct EBI-2 nucleotide and amino acid sequences would have been apparent to one skilled in the art in possession of ATCC Deposit No. 209003 and the data from the HHPGS02 sequencing run, as of the May 7, 1997 filing date of the present application. This is so because the correct EBI-2 coding sequence can be readily determined from the deposited clone and methods for sequencing this clone were routine in the art in May of 1997.

***Human Endothelium Differentiation Gene-1-Like (EDG-1-Like) G-Protein Coupled Receptor***

9. We obtained a cDNA clone encoding a human endothelium differentiation gene-1-like (EDG-1-like) G-protein coupled receptor by screening a cDNA library derived from human activated neutrophils. This clone was designated HNFDL69. We determined nucleotide sequence information for the HNFDL69 clone, as described below, using sequencing methods which were routine and publicly available as of the May 7, 1997 filing date of the present application. The



HNFDL69 clone that we obtained this sequence information from was deposited with the American Type Culture Collection (ATCC) on April 28, 1997 and was given ATCC Accession No. 209004 (*See* Attachment A.)

10. Evidence that the human HNFDL69 cDNA was deposited at the ATCC as Accession No. 209004 is provided by comparing the ATCC Deposit Receipt (Attachment A) with the information provided in the IRIS notebook page (entitled "Sequence Worksheet") included herewith as the first page of Attachment C. The section of the page entitled "Sequence Information" corresponds clone HNFDL69 with the "HGS Code" 563238. HGS Code 563238, represents a particular sequence entry in IRIS for cDNA clone, HNFDL69. HGS code 563238 appears as the identifier on the ATCC deposit receipt. (*See* Attachment A.) This indicates that the clone used to obtain the sequence information of HGS Code 563238 was deposited. In other words, even though, as explained below, SEQ ID NO:3 and SEQ ID NO:4 of the Sequence Listing in the present application as originally filed, had typographical errors due to attorney error, the human HNFDL69 cDNA clone used to obtain the original, correct sequence data was deposited at the ATCC.

11. Attachment C provides three pages of data from the IRIS electronic notebook which shows the sequencing results of the human EDG-1-like cDNA clone (*i.e.*, HNFDL69). The HNFDL69 sequence was obtained using a 373 Automated DNA sequencer (Applied Biosystems, Inc.). Sequencing accuracy using this method is predicted to be greater than 97%.

12. The information obtained from the HNFDL69 nucleotide sequencing run differs from the Sequence Listing currently on file in the present application in two positions. In particular, SEQ ID NO:3 contains typographical errors at the following nucleotide positions: position 828, which should be T rather than C; and position 831, which should be T rather than

A. Note that this latter typographical error introduced a stop codon into the open reading frame, causing the amino acid sequence, as translated from the sequence with the typographical error, to stop at position 260. Accordingly, these typographical errors in the nucleotide sequence result in the following errors to the encoded amino acid sequence depicted as SEQ IDNO:4: the serine at position 260 should be replaced with phenylalanine, and the translation should continue to amino acid 384, as depicted in the original translation provided on the third page of Attachment


C. SEQ ID NO:4 further contains typographical errors at the following amino acid positions: position 191, which should be Asp rather than Asn, position 202, which should be Lys rather than Arg, and position 204, which should be Tyr rather than Thr. In addition, the translation should start with the Met at position 1, rather than the Ala at position -16. Both of the nucleotide sequence changes are reflected in the HNFDL69 nucleotide sequence data shown on the first and second pages of Attachment C, and the amino acid sequence changes are reflected in the HNFDL69 amino acid sequence data shown on the third page of Attachment C. In addition, The amino acid sequence data is reflected in an amino acid alignment originally filed with the present application as Figure 4, except for five residues at the 3' end of the polypeptide. These latter five residues are not in the alignment simply because they did not align with the second sequence in Figure 4, i.e., SEQ ID NO:18.

13. I believe that the actual nucleotide sequence of the human HNFDL69 cDNA clone is the same as that originally entered in the IRIS notebook.

14. I am of the opinion that the correct EDG-1-like nucleotide and amino acid sequences would have been apparent to one skilled in the art in possession of ATCC Deposit No. 209004 and the data from the HNFDL69 sequencing run, as of the May 7, 1997 filing date of the present application. This is so because the correct EDG-1-like coding sequence can be readily

[illegible]

2/15/99  
Date

  
Steven M. Ruben

SEQUENCE LISTING

<110> Li, Yi

Ruben, Steven M.

<120> Human G-Protein Coupled Receptors

<130> 1488.1220002

<140>

<141>

<150> 08/852,824

<151> 1997-05-07

<160> 18

<170> PatentIn Ver. 2.0

<210> 1

<211> 2247

<212> DNA

<213> genomic

<220>

<221> CDS

<222> (226)..(1251)

<400> 1

gcacgaggaa cagaacactt tctcatgtcc agggtcagat tacaagagca ctcaagactt 60

tactgacgaa aactcaggaa atcctctatc acaaagaggt ttggcaacta aactaagaca 120

ttaaaaggaa aataccagat gccactctgc aggetgcaat aactactact tactggatac 180

attcaaacc tccagaatca acagttatca ggtaaccaac aagaa atg caa gcc gtc 237

Met Gln Ala Val

1

gac aat ctc acc tct gcg cct ggg aac acc agt ctg tgc acc aga gac 285

Asp Asn Leu Thr Ser Ala Pro Gly Asn Thr Ser Leu Cys Thr Arg Asp

5

10

15

20

tac aaa atc acc cag gtc ctc ttc cca ctg ctc tac act gtc ctg ttt 333  
Tyr Lys Ile Thr Gln Val Leu Phe Pro Leu Leu Tyr Thr Val Leu Phe  
25 30 35

ttt gtt gga ctt atc aca aat ggc ctg gcg atg agg att ttc ttt caa 381  
Phe Val Gly Leu Ile Thr Asn Gly Leu Ala Met Arg Ile Phe Phe Gln  
40 45 50

atc cgg agt aaa tca aac ttt att att ttt ctt aag aac aca gtc att 429  
Ile Arg Ser Lys Ser Asn Phe Ile Ile Phe Leu Lys Asn Thr Val Ile  
55 60 65

tct gat ctt ctc atg att ctg act ttt cca ttc aaa att ctt agt gat 477  
Ser Asp Leu Leu Met Ile Leu Thr Phe Pro Phe Lys Ile Leu Ser Asp  
70 75 80

gcc aaa ctg gga aca gga cca ctg aga act ttt gtg tgt caa gtt acc 525  
Ala Lys Leu Gly Thr Gly Pro Leu Arg Thr Phe Val Cys Gln Val Thr  
85 90 95 100

tcc gtc ata ttt tat ttc aca atg tat atc agt att tca ttc ctg gga 573  
Ser Val Ile Phe Tyr Phe Thr Met Tyr Ile Ser Ile Ser Phe Leu Gly  
105 110 115

ctg ata act atc gat cgc tac cag aag acc acc agg cca ttt aaa aca 621  
Leu Ile Thr Ile Asp Arg Tyr Gln Lys Thr Thr Arg Pro Phe Lys Thr  
120 125 130

tcc aac ccc aaa aat ctc ttg ggg gct aag att ctc tct gtt gtc atc 669  
Ser Asn Pro Lys Asn Leu Leu Gly Ala Lys Ile Leu Ser Val Val Ile  
135 140 145

tgg gca ttc atg ttc tta ctc tct ttg cct aac atg att ctg acc aac 717  
Trp Ala Phe Met Phe Leu Leu Ser Leu Pro Asn Met Ile Leu Thr Asn  
150 155 160

agg cag cgg aga gac aag aat gtg aag aaa tgc tct ttc ctt aaa tca 765  
Arg Gln Pro Arg Asp Lys Asn Val Lys Lys Cys Ser Phe Leu Lys Ser  
165 170 175 180

gag ttc ggt cta gtc tgg cat gaa ata gta aat tac atc tgt caa gtc	813
Glu Phe Gly Leu Val Trp His Glu Ile Val Asn Tyr Ile Cys Gln Val	
185 190 195	
att ttc tgg att aat ttc tta att gtt att gta tgt tat aca ctc att	861
Ile Phe Trp Ile Asn Phe Leu Ile Val Ile Val Cys Tyr Thr Leu Ile	
200 205 210	
aca aaa gaa ctg tac cgg tca tac gta aga acg agg ggt gta ggt aaa	909
Thr Lys Glu Leu Tyr Arg Ser Tyr Val Arg Thr Arg Gly Val Gly Lys	
215 220 225	
gtc ccc agg aaa aag gtg aac gtc aaa gtt ttc att atc att gct gta	957
Val Pro Arg Lys Lys Val Asn Val Lys Val Phe Ile Ile Ile Ala Val	
230 235 240	
ttc ttt att tgt ttt gtt cct ttc cat ttt gcc cga att cct tac acc	1005
Phe Phe Ile Cys Phe Val Pro Phe His Phe Ala Arg Ile Pro Tyr Thr	
245 250 255 260	
ctg agc caa acc cgg gat gtc ttt gac tgc act gct gaa aat act ctg	1053
Leu Ser Gln Thr Arg Asp Val Phe Asp Cys Thr Ala Glu Asn Thr Leu	
265 270 275	
ttc tat gtg aaa gag agc act ctg tgg tta act tcc tta aat gca tgc	1101
Phe Tyr Val Lys Glu Ser Thr Leu Trp Leu Thr Ser Leu Asn Ala Cys	
280 285 290	
ctg gat ccg ttc atc tat ttt ttc ctt tgc aag tcc ttc aga aat tcc	1149
Leu Asp Pro Phe Ile Tyr Phe Phe Leu Cys Lys Ser Phe Arg Asn Ser	
295 300 305	
ttg ata agt atg ctg aag tgc ccc aat tct gca aca tct ctg tcc cag	1197
Leu Ile Ser Met Leu Lys Cys Pro Asn Ser Ala Thr Ser Leu Ser Gln	
310 315 320	
gac aat agg aaa aaa gaa cag gat ggt ggt gac cca aat gaa gag act	1245
Asp Asn Arg Lys Lys Glu Gln Asp Gly Gly Asp Pro Asn Glu Glu Thr	
325 330 335 340	
cca atg taaacaaatt aactaaggaa atatttcaat ctctttgtgt tcagaactcg	1301

00515361 000000

Pro Met

ttaaagcaaa gcgctaagta aaaatattaa ctgacgaaga agcaactaag ttaataataa 1361  
 tgactctaaa gaaacagaag attacaaaag caattttcat ttacctttcc agtatgaaaa 1421  
 gctatcttaa aatatagaaa actaatctaa actgtagctg tattagcagc aaaacaaacg 1481  
 acatccaatt gtcattgctgc atgcaaaact acacagaatt catgttttgg cagagttttg 1541  
 gcaaaatgag taatcatata atatttactg taatttttaa aatacattat cgttcacaat 1601  
 tttatttttt cataatcaac taaggaagaa cgatcaattg gatataatct tcttaccaaa 1661  
 aatgatagtt aaaaagtata tatatcctag tcccctaacc aaatcctgac ctattgggat 1721  
 acttataaaa atttaagtaa gtgggataca caaagaataa taactattaa cttttcatta 1781  
 ttaggcaaaa acctaaggga tttaaaactaa ttgaaactgt atttgattgg acttaatttt 1841  
 ttatgtttat ttagaagata aagattttaag aagaccttta caataaagag aagaaatatc 1901  
 gaagtcatia aaataaggag acttactttt atgacattct aatactaaaa aatatagaaa 1961  
 tatttcotta attotagaga aactagtttt actaattttt tacaacttca ataataccat 2021  
 cactgacact tacctttatt aattagcttc tagaaaatag ctgctaatta ggttaatgaa 2081  
 cattttacct tagtgaaaaa aaattaatta aatatgatta caaagttgca cagcataact 2141  
 actgagagga aagtgattga tctgttttga attacttggt tgtattgggtg tgtataaaaa 2201  
 aaaaatttac attaaactct aaatcattaa aaaaaaaaaa aaaaaa 2248

<210> 2

<211> 342

<212> PRT

<213> genomic

<400> 2

Met Gln Ala Val Asp Asn Leu Thr Ser Ala Pro Gly Asn Thr Ser Leu  
1 5 10 15

Cys Thr Arg Asp Tyr Lys Ile Thr Gln Val Leu Phe Pro Leu Leu Tyr  
20 25 30

Thr Val Leu Phe Phe Val Gly Leu Ile Thr Asn Gly Leu Ala Met Arg  
35 40 45

Ile Phe Phe Gln Ile Arg Ser Lys Ser Asn Phe Ile Ile Phe Leu Lys  
50 55 60

Asn Thr Val Ile Ser Asp Leu Leu Met Ile Leu Thr Phe Pro Phe Lys  
65 70 75 80

Ile Leu Ser Asp Ala Lys Leu Gly Thr Gly Pro Leu Arg Thr Phe Val  
85 90 95

Cys Gln Val Thr Ser Val Ile Phe Tyr Phe Thr Met Tyr Ile Ser Ile  
100 105 110

Ser Phe Leu Gly Leu Ile Thr Ile Asp Arg Tyr Gln Lys Thr Thr Arg  
115 120 125

Pro Phe Lys Thr Ser Asn Pro Lys Asn Leu Leu Gly Ala Lys Ile Leu  
130 135 140

Ser Val Val Ile Trp Ala Phe Met Phe Leu Leu Ser Leu Pro Asn Met  
145 150 155 160

Ile Leu Thr Asn Arg Gln Pro Arg Asp Lys Asn Val Lys Lys Cys Ser  
165 170 175

Phe Leu Lys Ser Glu Phe Gly Leu Val Trp His Glu Ile Val Asn Tyr  
180 185 190

Ile Cys Gln Val Ile Phe Trp Ile Asn Phe Leu Ile Val Ile Val Cys  
195 200 205

Tyr Thr Leu Ile Thr Lys Glu Leu Tyr Arg Ser Tyr Val Arg Thr Arg  
210 215 220



Gly Val Gly Lys Val Pro Arg Lys Lys Val Asn Val Lys Val Phe Ile  
225 230 235 240

Ile Ile Ala Val Phe Phe Ile Cys Phe Val Pro Phe His Phe Ala Arg  
245 250 255

Ile Pro Tyr Thr Leu Ser Gln Thr Arg Asp Val Phe Asp Cys Thr Ala  
260 265 270

Glu Asn Thr Leu Phe Tyr Val Lys Glu Ser Thr Leu Trp Leu Thr Ser  
275 280 285

Leu Asn Ala Cys Leu Asp Pro Phe Ile Tyr Phe Phe Leu Cys Lys Ser  
290 295 300

Phe Arg Asn Ser Leu Ile Ser Met Leu Lys Cys Pro Asn Ser Ala Thr  
305 310 315 320

Ser Leu Ser Gln Asp Asn Arg Lys Lys Glu Gln Asp Gly Gly Asp Pro  
325 330 335

Asn Glu Glu Thr Pro Met  
340

<210> 3

<211> 1637

<212> DNA

<213> genomic

<220>

<221> CDS

<222> (50)..(1201)

<400> 3

ggcagcgacc caccctgcgt cgggcctcag tcagcccccg ggggaggcc atg aac gcc 58  
Met Asn Ala

1

acg ggg acc ccg gtg gcc ccc gag tcc tgc caa cag ctg gcg gcc ggc 106

Thr	Gly	Thr	Pro	Val	Ala	Ala	Pro	Glu	Ser	Cys	Gln	Gln	Leu	Ala	Ala	Gly	
5							10					15					
ggg	cac	agc	cgg	ctc	att	gtt	ctg	cac	tac	aac	cac	tcg	ggc	cgg	ctg		154
Gly	His	Ser	Arg	Leu	Ile	Val	Leu	His	Tyr	Asn	His	Ser	Gly	Arg	Leu		
20					25					30					35		
gcc	ggg	cgc	ggg	ggg	ccg	gag	gat	ggc	ggc	ctg	ggg	gcc	ctg	cgg	ggg		202
Ala	Gly	Arg	Gly	Gly	Pro	Glu	Asp	Gly	Gly	Leu	Gly	Ala	Leu	Arg	Gly		
				40					45					50			
ctg	tcg	gtg	gcc	gcc	agc	tgc	ctg	gtg	gtg	ctg	gag	aac	ttg	ctg	gtg		250
Leu	Ser	Val	Ala	Ala	Ser	Cys	Leu	Val	Val	Leu	Glu	Asn	Leu	Leu	Val		
			55					60					65				
ctg	gcg	gcc	atc	acc	agc	cac	atg	cgg	tcg	caa	cgc	tggtg	gtc	tac	tat		298
Leu	Ala	Ala	Ile	Thr	Ser	His	Met	Arg	Ser	Gln	Arg	Trp	Val	Tyr	Tyr		
	70						75					80					
tgc	ctg	gtg	aac	att	acg	atg	agt	gac	ctg	ctc	acg	ggc	gcg	gcc	tac		346
Cys	Leu	Val	Asn	Ile	Thr	Met	Ser	Asp	Leu	Leu	Thr	Gly	Ala	Ala	Tyr		
	85						90				95						
ctg	gcc	aac	gtg	ctg	ctg	tcg	ggg	gcc	cgc	acc	ttc	cgt	ctg	gcg	ccc		394
Leu	Ala	Asn	Val	Leu	Leu	Ser	Gly	Ala	Arg	Thr	Phe	Arg	Leu	Ala	Pro		
100				105					110					115			
gcc	cag	tgg	ttc	cta	cgg	aag	ggc	ctg	ctc	ttc	acc	gcc	ctg	gcc	gcc		442
Ala	Gln	Trp	Phe	Leu	Arg	Lys	Gly	Leu	Leu	Phe	Thr	Ala	Leu	Ala	Ala		
				120					125					130			
tcc	acc	ttc	agc	ctg	ctc	ttc	act	gca	ggg	ttg	cgc	ttt	gcc	acc	atg		490
Ser	Thr	Phe	Ser	Leu	Leu	Phe	Thr	Ala	Gly	Leu	Arg	Phe	Ala	Thr	Met		
			135					140					145				
gtg	cgg	ccg	gtg	gcc	gag	agc	ggg	gcc	acc	aag	acc	agc	cgc	gtc	tac		538
Val	Arg	Pro	Val	Ala	Glu	Ser	Gly	Ala	Thr	Lys	Thr	Ser	Arg	Val	Tyr		
	150						155					160					
ggc	ttc	atc	ggc	ctc	tgc	tgg	ctg	ctg	gcc	gcg	ctg	ctg	ggg	atg	ctg		586
Gly	Phe	Ile	Gly	Leu	Cys	Trp	Leu	Leu	Ala	Ala	Leu	Leu	Gly	Met	Leu		

165	170	175	
cct ttg ctg ggc tgg aac tgc ctg tgc gcc ttt gac cgc tgc tcc agc			634
Pro Leu Leu Gly Trp Asn Cys Leu Cys Ala Phe Asp Arg Cys Ser Ser			
180	185	190	195
ctt ctg ccc ctc tac tcc aag cgc tac atc ctc ttc tgc ctg gtg atc			682
Leu Leu Pro Leu Tyr Ser Lys Arg Tyr Ile Leu Phe Cys Leu Val Ile			
200	205		210
ttc gcc ggc gtc ctg gcc acc atc atg ggc ctc tat ggg gcc atc ttc			730
Phe Ala Gly Val Leu Ala Thr Ile Met Gly Leu Tyr Gly Ala Ile Phe			
215	220		225
cgc ctg gtg cag gcc agc ggg cag aag gcc cca cgc cca gcg gcc cgc			778
Arg Leu Val Gln Ala Ser Gly Gln Lys Ala Pro Arg Pro Ala Ala Arg			
230	235		240
cgc aag gcc cgc cgc ctg ctg aag acg gtg ctg atg atc ctg ctg gcc			826
Arg Lys Ala Arg Arg Leu Leu Lys Thr Val Leu Met Ile Leu Leu Ala			
245	250		255
ttc ttg gtg tgc tgg gga cca ctc ttc ggg ctg ctg ctg gcc gac gtc			874
Phe Leu Val Cys Trp Gly Pro Leu Phe Gly Leu Leu Leu Ala Asp Val			
260	265	270	275
ttt ggc tcc aac ctc tgg gcc cag gag tac ctg cgg gcc atg gac tgg			922
Phe Gly Ser Asn Leu Trp Ala Gln Glu Tyr Leu Arg Gly Met Asp Trp			
280	285		290
atc ctg gcc ctg gcc gtc ctc aac tgc gcg gtc aac ccc atc atc tac			970
Ile Leu Ala Leu Ala Val Leu Asn Ser Ala Val Asn Pro Ile Ile Tyr			
295	300		305
tcc ttc cgc agc agg gag gtg tgc aga gcc gtg ctc agc ttc ctc tgc			1018
Ser Phe Arg Ser Arg Glu Val Cys Arg Ala Val Leu Ser Phe Leu Cys			
310	315		320
tgc ggg tgt ctc cgg ctg ggc atg cga ggg ccc ggg gac tgc ctg gcc			1066
Cys Gly Cys Leu Arg Leu Gly Met Arg Gly Pro Gly Asp Cys Leu Ala			
325	330		335

cgg gcc gtc gag gct cac tcc gga gct tcc acc acc gac agc tct ctg 1114  
Arg Ala Val Glu Ala His Ser Gly Ala Ser Thr Thr Asp Ser Ser Leu  
340 345 350 355

agg cca agg gac agc ttt cgc ggc tcc cgc tcg ctc agc ttt cgg atg 1162  
Arg Pro Arg Asp Ser Phe Arg Gly Ser Arg Ser Leu Ser Phe Arg Met  
360 365 370

cgg gag ccc ctg tcc agc atc tcc agc gtg cgg agc atc tgaagttgca 1211  
Arg Glu Pro Leu Ser Ser Ile Ser Ser Val Arg Ser Ile  
375 380

gtcttgcgtg tggatggtgc aaccacggg tgcgtgccag gcaggccctc ctgggggtaca 1271

ggaagctgtg tgcacgcaac ctgcacctgt atggggagca gggaacggga caggccccc 1331

tggacttgcc cgggtggcctc toggggcttc tgacgccata tggacttgcc cattgcctat 1391

ggctcacctt ggacaaggag gcaaccaccc cacctccccg taggagcaga gagcaccctg 1451

gtgtgggggc gagtgggttc ccacacaacc cgcttctgtg tgattctggg gaagtcccg 1511

cccctctctg ggctcagta gggctcccag gctgcaaggg gtggactgtg ggatgcatgc 1571

cctggcaaca ttgaagtctg atcatggtaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1631

aaaaaa 1637

<210> 4

<211> 384

<212> PRT

<213> genomic

<400> 4

Met Asn Ala Thr Gly Thr Pro Val Ala Pro Glu Ser Cys Gln Gln Leu  
1 5 10 15

Ala Ala Gly Gly His Ser Arg Leu Ile Val Leu His Tyr Asn His Ser  
20 25 30

Gly Arg Leu Ala Gly Arg Gly Gly Pro Glu Asp Gly Gly Leu Gly Ala  
35 40 45

Leu Arg Gly Leu Ser Val Ala Ala Ser Cys Leu Val Val Leu Glu Asn  
50 55 60

Leu Leu Val Leu Ala Ala Ile Thr Ser His Met Arg Ser Gln Arg Trp  
65 70 75 80

Val Tyr Tyr Cys Leu Val Asn Ile Thr Met Ser Asp Leu Leu Thr Gly  
85 90 95

Ala Ala Tyr Leu Ala Asn Val Leu Leu Ser Gly Ala Arg Thr Phe Arg  
100 105 110

Leu Ala Pro Ala Gln Trp Phe Leu Arg Lys Gly Leu Leu Phe Thr Ala  
115 120 125

Leu Ala Ala Ser Thr Phe Ser Leu Leu Phe Thr Ala Gly Leu Arg Phe  
130 135 140

Ala Thr Met Val Arg Pro Val Ala Glu Ser Gly Ala Thr Lys Thr Ser  
145 150 155 160

Arg Val Tyr Gly Phe Ile Gly Leu Cys Trp Leu Leu Ala Ala Leu Leu  
165 170 175

Gly Met Leu Pro Leu Leu Gly Trp Asn Cys Leu Cys Ala Phe Asp Arg  
180 185 190

Cys Ser Ser Leu Leu Pro Leu Tyr Ser Lys Arg Tyr Ile Leu Phe Cys  
195 200 205

Leu Val Ile Phe Ala Gly Val Leu Ala Thr Ile Met Gly Leu Tyr Gly  
210 215 220

Ala Ile Phe Arg Leu Val Gln Ala Ser Gly Gln Lys Ala Pro Arg Pro  
225 230 235 240

Ala Ala Arg Arg Lys Ala Arg Arg Leu Leu Lys Thr Val Leu Met Ile

-11-

245

250

255

Leu Leu Ala Phe Leu Val Cys Trp Gly Pro Leu Phe Gly Leu Leu Leu  
260 265 270

Ala Asp Val Phe Gly Ser Asn Leu Trp Ala Gln Glu Tyr Leu Arg Gly  
275 280 285

Met Asp Trp Ile Leu Ala Leu Ala Val Leu Asn Ser Ala Val Asn Pro  
290 295 300

Ile Ile Tyr Ser Phe Arg Ser Arg Glu Val Cys Arg Ala Val Leu Ser  
305 310 315 320

Phe Leu Cys Cys Gly Cys Leu Arg Leu Gly Met Arg Gly Pro Gly Asp  
325 330 335

Cys Leu Ala Arg Ala Val Glu Ala His Ser Gly Ala Ser Thr Thr Asp  
340 345 350

Ser Ser Leu Arg Pro Arg Asp Ser Phe Arg Gly Ser Arg Ser Leu Ser  
355 360 365

Phe Arg Met Arg Glu Pro Leu Ser Ser Ile Ser Ser Val Arg Ser Ile  
370 375 380

<210> 5

<211> 28

<212> DNA

<213> genomic

<400> 5

ccgaggatcc atgcaagccg tcgacaat

28

<210> 6

<211> 28

<212> DNA

<213> genomic

<400> 6  
ccgaggatcc ttacattgga gtctcttc 28

<210> 7  
<211> 34  
<212> DNA  
<213> genomic

<400> 7  
ccgaggatcc gccatcatgc aagccgtcga caat 34

<210> 8  
<211> 28  
<212> DNA  
<213> genomic

<400> 8  
ccgaggatcc ttacattgga gtctcttc 28

<210> 9  
<211> 34  
<212> DNA  
<213> genomic

<400> 9  
ccgaggatcc gccatcatgc aagccgtcga caat 34

<210> 10  
<211> 55  
<212> DNA  
<213> genomic

<400> 10  
ccgatctaga ttaatcccat acgacgtccc agactacgct cattggagtc tcttc 55

<210> 11

<211> 28  
<212> DNA  
<213> genomic

<400> 11  
ccgaggatcc atgaacgccg cgggggacc 28

<210> 12  
<211> 28  
<212> DNA  
<213> genomic

<400> 12  
ccgaggatcc tcagatgctc cgcacgct 28

<210> 13  
<211> 34  
<212> DNA  
<213> genomic

<400> 13  
gcgaggatcc gccatcatga acgccacggg gacc 34

<210> 14  
<211> 28  
<212> DNA  
<213> genomic

<400> 14  
ccgaggatcc tcagatgctc cgcacgct 28

<210> 15  
<211> 34  
<212> DNA  
<213> genomic

<400> 15

CGAGGATCC ATGAACGCCG CGGGGACC



ccgaggatcc gccatcatga acgccacggg gacc

34

<210> 16

<211> 55

<212> DNA

<213> genomic

<400> 16

ccgatctaga tcaatcccat acgacgtccc agactacgct gatgctccg acgct

55

<210> 17

<211> 348

<212> PRT

<213> genomic

<400> 17

Ile Gln Met Ala Asn Asn Phe Thr Pro Pro Ser Ala Thr Pro Gln Asn

1

5

10

15

Asp Cys Asp Leu Tyr Ala His His Ser Thr Ala Arg Ile Val Met Pro

20

25

30

Leu His Tyr Ser Leu Val Phe Ile Ile Gly Leu Val Gly Asn Leu Leu

35

40

45

Ala Leu Val Val Ile Val Gln Asn Arg Lys Lys Ile Asn Ser Thr Thr

50

55

60

Leu Tyr Ser Thr Asn Leu Val Ile Ser Asp Ile Leu Phe Thr Thr Ala

65

70

75

80

Leu Pro Thr Arg Ile Ala Tyr Tyr Ala Met Gly Phe Asp Trp Arg Ile

85

90

95

Gly Asp Ala Leu Cys Arg Ile Thr Ala Leu Val Phe Tyr Ile Asn Thr

100

105

110

Tyr Ala Gly Val Asn Phe Met Thr Cys Leu Ser Ile Asp Arg Phe Ile

115

120

125

Ala Val Val His Pro Leu Arg Tyr Asn Lys Ile Lys Arg Ile Glu His  
130 135 140

Ala Lys Gly Val Cys Ile Phe Val Trp Ile Leu Val Phe Ala Gln Thr  
145 150 155 160

Leu Pro Leu Leu Ile Asn Pro Met Ser Lys Gln Glu Ala Glu Arg Ile  
165 170 175

Thr Cys Met Glu Tyr Pro Asn Phe Glu Glu Thr Lys Ser Leu Pro Trp  
180 185 190

Ile Leu Leu Gly Ala Cys Phe Ile Gly Tyr Val Leu Pro Leu Ile Ile  
195 200 205

Ile Lys Ile Cys Tyr Ser Gln Ile Cys Cys Lys Leu Phe Arg Thr Ala  
210 215 220

Lys Gln Asn Pro Leu Thr Glu Lys Ser Gly Val Asn Lys Lys Ala Leu  
225 230 235 240

Asn Thr Ile Ile Leu Ile Ile Val Val Phe Val Leu Cys Phe Thr Pro  
245 250 255

Tyr His Val Ala Ile Ile Gln His Met Ile Lys Lys Leu Arg Phe Ser  
260 265 270

Asn Phe Leu Glu Cys Ser Gln Arg His Ser Phe Gln Ile Ser Leu His  
275 280 285

Phe Thr Val Cys Leu Met Asn Phe Asn Cys Cys Met Asp Pro Phe Ile  
290 295 300

Tyr Phe Phe Ala Cys Lys Gly Tyr Lys Arg Lys Val Met Arg Met Leu  
305 310 315 320

Lys Arg Gln Val Ser Val Ser Ile Ser Ser Ala Val Lys Ser Ala Pro  
325 330 335

Glu Glu Asn Ser Arg Glu Met Thr Glu Thr Gln Met

340

345

<210> 18

<211> 381

<212> PRT

<213> genomic

<400> 18

Met Gly Pro Thr Ser Val Pro Leu Val Lys Ala His Arg Ser Ser Val

1

5

10

15

Ser Asp Tyr Val Asn Tyr Asp Ile Ile Val Arg His Tyr Asn Tyr Thr

20

25

30

Gly Lys Leu Asn Ile Ser Ala Asp Lys Glu Asn Ser Ile Lys Leu Thr

35

40

45

Ser Val Val Phe Ile Leu Ile Cys Cys Phe Ile Ile Leu Glu Asn Ile

50

55

60

Phe Val Leu Leu Thr Ile Trp Lys Thr Lys Lys Phe His Arg Pro Met

65

70

75

80

Tyr Tyr Phe Ile Gly Asn Leu Ala Leu Ser Asp Leu Leu Ala Gly Val

85

90

95

Ala Tyr Thr Ala Asn Leu Leu Leu Ser Gly Ala Thr Thr Tyr Lys Leu

100

105

110

Thr Pro Ala Gln Trp Phe Leu Arg Glu Gly Ser Met Phe Val Ala Leu

115

120

125

Ser Ala Ser Val Phe Ser Leu Leu Ala Ile Ala Ile Glu Arg Tyr Ile

130

135

140

Thr Met Leu Lys Met Lys Leu His Asn Gly Ser Asn Asn Phe Arg Leu

145

150

155

160

Phe Leu Leu Ile Ser Ala Cys Trp Val Ile Ser Leu Ile Leu Gly Gly

165

170

175

Glu Thr Ile Met Ser Ser Gly Asn Val Asn Ser Ser Ser  
370 375 380